Polypeptide nucleic sequences exported from mycobacteria, vectors comprising same and uses for diagnosing and preventing tuberculosis

The subject of the invention is novel recombinant screening, cloning and/or expression vectors which replicate in mycobacteria. Its subject is also a set of sequences encoding exported polypeptides which are detected by fusions with alkaline phosphatase and whose expression is regulated (induced or repressed) or constitutive during the ingestion of mycobacteria by macrophages. The invention also relates to a polypeptide, called DP428, of about 12 kD which corresponds to an exported protein found in mycobacteria belonging to the *Mycobacterium tuberculosis* complex. The invention also relates to a polynucleotide comprising a sequence encoding this polypeptide. It also relates to the use of the polypeptide or of fragments thereof and of the polynucleotides encoding the latter (or alternatively the polynucleotides complementary to the latter) for the production of means for detecting in vitro or in vivo the presence of a mycobacterium belonging to the *Mycobacterium tuberculosis* complex in a biological sample or for the detection of reactions of the host infected with these bacterial species. The invention finally relates to the use of the polypeptide or of fragments thereof as well as of the polynucleotides encoding the latter as means intended for the preparation of an immunogenic composition which is capable of inducing an immune response directed against the mycobacteria belonging to the *Mycobacterium tuberculosis* complex, or of a vaccine composition for the prevention and/or treatment of infections caused by mycobacteria belonging to said complex, in particular tuberculosis.

The aim of the present invention is also to use these sequences (polypeptide and polynucleotide

sequences) as target for the search for novel inhibitors of the growth and multiplication of mycobacteria and of their maintenance in the host, it being possible for these inhibitors to serve as antibiotics.

The genus Mycobacterium, which comprises at least 56 different species, includes major human pathogens such as *M. leprae* and *M. tuberculosis*, the agents responsible for leprosy and tuberculosis, which remain serious public health problems worldwide.

Tuberculosis continues to be a public health problem in the world. At present, this disease is the cause of 2 to 3 million deaths in the world and about 8 million new cases are observed each year (Bouvet, 1994). In developed countries, *M. tuberculosis* is the most common cause of mycobacteria infections. In France, about 10,000 new cases appear per year and, among the notifiable diseases, it is tuberculosis which comprises the highest number of cases. Vaccination with BCG (Bacille Calmette-Guérin), an avirulent strain which is derived from *M. bovis* and which is widely used as a vaccine against tuberculosis, is far from being effective in all populations. This efficacy varies from about 80% in western countries such as England, to 0% in India (results of the last vaccination trial in Chingleput., published in 1972 in Indian J. Med. Res.). Furthermore, the appearance of *M. tuberculosis* strains which are resistant to antituberculars and the increased risk in immunosuppressed patients, patients suffering from AIDS, of developing tuberculosis, make the development of rapid, specific and reliable methods for the diagnosis of tuberculosis and the development of novel vaccines necessary. For example, an epidemiological study carried out in Florida, and of which the results were published in 1993 in AIDS therapies, showed that 10% of the AIDS patients are affected by tuberculosis at the time of the AIDS diagnosis or 18 months before it. In these patients, tuberculosis appears in 60% of cases in a form which is

disseminated and therefore nondetectable by conventional diagnostic criteria such as pulmonary radiography or the analysis of sputum.

Currently, a certainty on the diagnosis provided by the detection of bacilli which can be cultured in a sample obtained from a patient is obtained in only less than half of the tuberculosis cases, even in the case of pulmonary tuberculosis. The diagnosis of tuberculosis and of the other related mycobacteria is therefore difficult to carry out for various reasons: mycobacteria are often present in a small quantity, their generation time is very long (24 h for *M. tuberculosis*) and they are difficult to culture (Bates et al., 1986).

Other techniques can be used in clinical medicine to identify a mycobacterial infection:

a) The direct identification of microorganisms under a microscope; this technique is rapid, but does not allow the identification of the mycobacterial species observed and lacks sensitivity (Bates, 1979).

Cultures, when they are positive, have a specificity approaching 100% and allow the identification of the mycobacterial species isolated; however, as specified above, the growth of mycobacteria *in vitro* is long (can only be carried out in 3 to 6 weeks of repeated cultures (Bates, 1979; Bates et al., 1986)) and expensive.

b) Serological techniques are found to be useful under certain conditions, but their use is sometimes limited by their low sensitivity and/or specificity (Daniel et al., 1987).

c) The presence of mycobacteria in a biological sample can also be determined by molecular hybridization with DNA or RNA using oligonucleotide probes which are specific for the sequences tested for (Kiehn et al., 1987; Roberts et al., 1987; Drake et al., 1987). Several studies have shown the advantage of this technique for the diagnosis of mycobacterial infections. The probes used consist of DNA, ribosomal

RNA or DNA fragments from mycobacteria which are obtained from gene banks. The principle of these techniques is based on the polymorphism of the nucleotide sequences of the fragments used or on the polymorphism of the adjacent regions. In all cases, they require the use of cultures and are not directly applicable to biological samples.

The low quantity of mycobacteria present in a biological sample and consequently the low quantity of target DNA to be detected in this sample can require the use of a specific amplification *in vitro* of the target DNA before its detection with the aid of the nucleotide probe and using *in vitro* amplification techniques such as PCR (polymerase chain reaction). The specific amplification of the DNA by the PCR technique can constitute the first stage of a method for detecting the presence of a mycobacterial DNA in a biological sample, the actual detection of the amplified DNA being carried out in a second stage with the aid of an oligonucleotide probe capable of specifically hybridizing with the amplified DNA.

A test for the detection of mycobacteria belonging to the *Mycobacterium tuberculosis* complex, by sandwich hybridization (test using a capture probe and a detection probe) was described by Chevrier et al. in 1993. The *Mycobacterium tuberculosis* complex is a group of mycobacteria which comprises *M. bovis-BCG, M. bovis, M. tuberculosis, M. africanum* and *M. microti*.

A method for the detection of low quantities of mycobacteria, belonging to the tuberculosis complex, by gene amplification and direct hybridization on biological samples has been developed. Said method uses the insertion sequence IS*6110* (European Patent EP 0,490,951 B1). Thierry et al. described in 1990 a sequence which is specific to the *Mycobacterium tuberculosis* complex and which is called IS6110. Some authors have proposed specifically amplifying the DNA obtained from *Mycobacterium* using nucleic primers in an amplification method, such as the polymerase chain

reaction (PCR). Patel et al. described in 1990 the use of several nucleic primers chosen from a sequence known as a probe in the identification of *M. tuberculosis*. However, the length of the fragments obtained using these primers was different from the expected theoretical length and several fragments of variable size were obtained. Furthermore, the authors observed the absence of hybridization of the amplified products with the plasmid which served to determine the primers. These results indicate that these primers might not be appropriate in the detection of the presence of *M. tuberculosis* in a biological sample and confirm the critical nature of the choice of the primers. The same year, J.L. Guesdon and D. Thierry described a method for the detection of *M. tuberculosis*, having a high sensitivity, by amplification of an *M. tuberculosis* DNA fragment located within the IS6110 sequence (European Patent EP 461,045) with the aid of primers generating amplified DNA fragments of constant length, even when the choice of the primers led to the amplification of long fragments (of the order of 1000 to 1500 bases) where the risk of interruption of the polymerization is high because of the effects of the secondary structure of the sequence. Other primers specific for the IS6110 sequence are described in European Patent No. EP-0,490,951.

The inventors have shown (unpublished results) that some clinical isolates of *Mycobacterium tuberculosis* lacked the insertion sequence IS6110 and could therefore not be detected with the aid of oligonucleotides specific for this sequence which could thus lead to false-negative diagnostic results. These results confirm a similar observation made by Yuen et al. in 1993. The impossibility of detecting these pathogenic strains which are potentially present in a biological sample collected from a patient is thus likely to lead to diagnostic difficulties or even to diagnostic errors. The availability of several sequences specific for the tubercule bacillus, within

which primers appropriate for amplification will be chosen, is important. The DP428 sequence described here may be used.

M. bovis and *M. tuberculosis*, the causative agents of tuberculosis, are facultative intracellular bacteria.

These agents have developed mechanisms to ensure their survival and their replication inside macrophage, one of the cell types which is supposed to eradicate invasion by microorganisms. These agents are capable of modulating the normal development of their phagosome and of preventing them from becoming differentiated into an acidic compartment rich in hydrolase (Clemens, 1979; Clemens et al., 1996; Sturgill-Koszycki et al., 1994 and Xu et al., 1994). However, this modulation is only possible if the bacterium is alive inside the phagosome, suggesting that compounds which are actively synthesized and/or secreted inside the cell are part of this mechanism. Exported proteins are probably involved in this mechanism. Despite major health problems linked to these pathogenic organisms, little is known on their exported and/or secreted proteins. SDS-PAGE analyses of *M. tuberculosis* culture filtrate show at least 30 secreted proteins (Altschul et al., 1990; Nagal et al., 1991 and Young et al., 1992). Some of them have been characterized, their genes cloned and sequenced (Borremans et al., 1989; Wiker et al., 1992 and Yamaguchi et al., 1989). Others, although being immunodominant antigens of major importance for inducing a protective immunity (Anderson et al., 1991 and Orme et al., 1993), have not been completely identified. In addition, it is probable that many exported proteins remain attached to the cell membrane and are consequently not present in the culture supernatants. It has been shown that the proteins located at the outer surface of various pathogenic bacteria, such as the 103 kDa invasin of *Yersina Pseudotuberculosis* (Isberg et al., 1987) or the 80 kDa

internalin of *Listeria monocytogenes* (Gaillard et al., 1991 and Dramsi et al., 1997) play an important role in the interactions with the host cells and, consequently, in the pathogenicity as well as in the induction of protective responses. Thus, a protein which is bound to the membrane would be important for the *M. tuberculosis* infection as well as for the induction of a protective response against this infection. These proteins could certainly be of interest for the preparation of vaccines.

Recently, the adaptation, to mycobacteria, of a genetic methodology for the identification and the phenotypic selection of export proteins has been described (Lim et al., 1995). This method uses *E. coli* periplasmic alkaline phosphatase (PhoA). A plasmid vector was constructed which allows the fusion of genes between a truncated *PhoA* gene and genes encoding exported proteins (Manoil et al., 1990).

Using this method, it has been possible to identify an *M. tuberculosis* gene (*erp* (Berthet et al., 1995)) exhibiting homologies with a 28 kDa exported protein of *M. leprae*, which is a frequent target of humoral responses of the lepromatous form of leprosy. A protein having amino acid motifs which are characteristic of plant desaturase (*des*) has also been characterized by the technique of fusion with PhoA.

However, this genetic method for identifying exported proteins does not make it possible to easily evaluate the intracellular expression of the corresponding genes. Such an evaluation is of crucial importance both for selecting good candidate vaccines and for understanding the interactions between bacteria and their host cells. The induction of the expression of virulence factor through pathogenic target cell contact has been described. It is the case, for example, for the *Yersinia pseudotuberculosis* Yops virulence factors (Petersson et al., 1996). *Shigella*, upon contact with the target cells, releases the Ipa

proteïns into the culture medium, and *Salmonella* synthesizes novel surface structures.

Taking into account the preceding text, a great need currently exists for developing novel vaccines against pathogenic microbacteria as well as novel specific, reliable and rapid diagnostic tests. These developments require the designing of even more efficient specific tools which make it possible, on the one hand, to isolate or to obtain sequences of novel specific, in particular immunogenic, polypeptides, and, on the other hand, to better understand the mechanism of the interactions between bacteria and their host cells such as in particular the induction of the expression of virulence factor. This is precisely the object of the present invention.

The inventors have defined and produced, for this purpose, novel vectors allowing the screening, cloning and/or expression of mycobacterial DNA sequences so as to identify, among these sequences, nucleic acids encoding proteins of interest, preferably exported proteins, which may be located on the bacterial membrane, and/or secreted proteins, and to identify among these sequences those which are induced or repressed during infection (intracellular growth).

Description

The present invention describes the use of the reporter gene *phoA* in mycobacteria. It makes it possible to identify systems for expression and export in a mycobacterial context. Many genes are only expressed in such a context, which shows the advantage of the present invention. During the cloning of DNA segments of strains of the *M. tuberculosis* complex fused with *phoA* into another mycobacterium such as *M. smegmatis*, the beginning of the gene, its regulatory regions and its regulator will be cloned, which will make it possible to observe a regulation. If this regulation is positive, the cloning of the regulator

will constitute an advantage for observing the expression and the export.

In the context of the invention, mycobacterium is understood to mean all the mycobacteria belonging to the various species listed by Wayne L.G. and Kubica G.P. (1980). Family Mycobacteriaceae in Bergey's manual of systematic bacteriology, J.P. Butler Ed. (Baltimore USA: Williams and Wilkins P. 1436-1457).

In some cases, the cloned genes are subjected in their original host to a negative regulation which makes the observation of the expression and of the export difficult in the original host. In this case, the cloning of the gene in the absence of its negative regulator, into a host not containing it, will constitute an advantage.

The invention also relates to novel mycobacterial polypeptides and to novel mycobacterial polynucleotides which may have been isolated by means of the preceding vectors and which are capable of entering into the preparation of compositions for the detection of a microbacterial infection, or for the protection against an infection caused by mycobacteria or for the search for inhibitors as is described above for DP428.

The subject of the invention is therefore a recombinant screening, cloning and/or expression vector, characterized in that it replicates in mycobacteria and in that it contains:

1) a replicon which is functional in mycobacteria;
2) a selectable marker;
3) a reporter cassette comprising:
   a) a multiple cloning site (polylinker),
   b) optionally a transcription terminator which is active in mycobacteria, upstream of the polylinker,
   c) a coding nucleotide sequence which is derived from a gene encoding a protein expression, export and/or secretion marker, said nucleotide sequence lacking its initiation codon and its regulatory sequences, and

d) a coding nucleotide sequence derived from a gene encoding a marker for the activity of promoters which are contained in the same fragment, said nucleotide sequence lacking its initiation codon. Optionally, the recombinant vector also contains a replicon which is functional in *E. coli*.

Preferably, the export and/or secretion marker is placed in the same orientation as the promoter activity marker.

Preferably, the recombinant screening vector according to the invention comprises, in addition, a transcription terminator placed downstream of the promoter activity marker, which is likely to allow the production of short transcripts which are found to be more stable and which consequently allow a higher level of expression of the products of translation.

The export and/or secretion marker is a nucleotide sequence whose expression, followed by export and/or secretion, depends on the regulatory elements which control its expression.

"Sequences or elements for regulating the expression of the production of polypeptides and its location" is understood to mean a transcriptional promoter sequence, a sequence comprising the ribosome-binding site (RBS), the sequences responsible for export and/or secretion such as the sequence termed signal sequence.

A first advantageous export and/or expression marker is a coding sequence derived from the *phoA* gene. Where appropriate, it is truncated such that the alkaline phosphatase activity is nevertheless capable of being restored when the truncated coding sequence is placed under the control of a promoter and of appropriate regulatory elements.

Other exposure, export and/or secretion markers may be used. There may be mentioned, by way of examples, a sequence of the gene for β-agarase, for the nuclease of a staphylococcus or for a β-lactamase.

Among the advantageous markers for the activity of promoters which are contained in the same fragment, a coding sequence derived from the firefly luciferase *luc* gene, provided with its initiation codon, is preferred.

Other markers for the activity of promoters which are contained in the same fragment may be used. There may be mentioned, by way of examples, a sequence of the gene for GFP (Green Fluorescent Protein).

The transcription terminator should be functional in mycobacteria. An advantageous terminator is, in this regard, the T4 coliphage terminator (tT4). Other appropriate terminators for carrying out the invention may be isolated using the technique presented in the examples, for example by means of an "omega" cassette (Prentki et al., 1984).

A vector which is particularly preferred for carrying out the invention is a plasmid chosen from the following plasmids which have been deposited at the CNCM (Collection Nationale de Cultures de Microorganismes, 25 rue de Docteur Roux, 75724 Paris cedex 15, France):

a) pJVEDa which was deposited at the CNCM under the No. I-1797, on 12/12/1996,

b) pJVEDb which was deposited at the CNCM under the No. I-1906, on 25 July 1997,

c) pJVEDc which was deposited at the CNCM under the No. I-1799, on 12/12/1996.

For the selection or the identification of mycobacterial nucleic acid sequences encoding polypeptides which are capable of being incorporated into immunogenic or antigenic compositions for the detection of an infection, or which are capable of inducing or repressing a mycobacterial virulence factor, the vector of the invention will comprise, at one of the multiple cloning sites of the polylinker, a nucleotide sequence of a mycobacterium in which the detection is carried out of the presence of sequences corresponding to exported and/or secreted polypeptides

which may be induced or repressed during the infection, or alternatively expressed or produced constitutively, their associated promoter and/or regulatory sequences which are capable of allowing or promoting the export and/or the secretion of said polypeptides of interest, or all or part of the genes of interest encoding said polypeptides.

Preferably, this sequence is obtained by physical fragmentation or by enzymatic digestion of the genomic DNA or of the DNA which is complementary to an RNA of a mycobacterium, preferably *M. tuberculosis* or chosen from *M. africanum, M. bovis, M. avium* or *M. leprae*.

The vectors of the invention may indeed also be used to determine the presence of sequences of interest, preferably corresponding to exported and/or secreted proteins, and/or capable of being induced or repressed or produced constitutively during the infection, in particular during phagocytosis by the macrophages, and, according to what was previously disclosed, in mycobacteria such as *M. africanum, M. bovis, M. avium* or *M. leprae* whose DNA or cDNA will have been treated by physical fragmentation or with defined enzymes.

According to a first embodiment of the invention, the enzymatic digesion of the genomic DNA or of the complementary DNA is carried out using *M. tuberculosis*.

Preferably, this DNA is digested with an enzyme such as sau3A, BclI or BglII.

Other digestive enzymes such as ScaI, ApaI, SacII or KpnI or alternatively nucleases or polymerases can naturally be used as long as they allow the production of fragments whose ends can be inserted into one of the cloning sites of the polylinker of the vector of the invention.

Where appropriate, the digestions with various enzymes will be carried out simultaneously.

Recombinant vectors which are preferred for carrying out the invention are chosen from the following recombinant vectors which have been deposited at the CNCM:

a) p6D7 which was deposited on 28 January 1997 at the CNCM under the No. I-1814,

b) p5A3 which was deposited on 28 January 1997 at the CNCM under the No. I-1815,

c) p5F6 which was deposited on 28 January 1997 at the CNCM under the No. I-1816,

d) p2A29 which was deposited on 28 January 1997 at the CNCM under the No. I-1817,

e) pDP428 which was deposited on 28 January 1997 at the CNCM under the No. I-1818,

f) p5B5 which was deposited on 28 January 1997 at the CNCM under the No. I-1819,

g) p1C7 which was deposited on 28 January 1997 at the CNCM under the No. I-1820,

h) p2D7 which was deposited on 28 January 1997 at the CNCM under the No. I-1821,

i) p1B7 which was deposited on 31 January 1997 at the CNCM under the No. I-1843,

j) pJVED/*M. tuberculosis* which was deposited on 25 July 1997 at the CNCM under the No. I-1907,

k) pM1C25 which was deposited on 4 August 1998 at the CNCM under the No. I-2062.

Among those which are most preferred, the recombinant vector pDP428 which was deposited on 28 January 1997 at the CNCM under the No. I-1818, and the vector pM1C25 which was deposited on 4 August 1998 at the CNCM under the No. I-2062 are preferred.

The subject of the invention is also a method of screening nucleotide sequences derived from mycobacteria in order to determine the presence of sequences corresponding to exported and/or secreted polypeptides which may be induced or repressed during the infection, their associated promoter and/or regulatory sequences which are capable in particular of allowing or promoting the export and/or secretion of

said polypeptides of interest, or all or part of genes of interest encoding said polypeptides, characterized in that it uses a recombinant vector according to the invention.

The invention also relates to a method of screening, according to the invention, characterized in that it comprises the following steps:

a) physical fragmentation of the mycobacterial DNA sequences or their digestion with at least one defined enzyme and recovery of the fragments obtained;

b) insertion of the fragments obtained in step a) into a cloning site, which is compatible, where appropriate, with the enzyme of step a), of the polylinker of a vector according to the invention;

c) if necessary, amplification of said fragments contained in the vector, for example by replication of the latter after insertion of the vector thus modified into a defined cell, preferably *E. coli*;

d) transformation of the host cells with the vector amplified in step c), or in the absence of amplification, with the vector of step b);

e) culture of the transformed host cells in a medium allowing the detection of the export and/or secretion marker, and/or of the promoter activity marker which is contained in the vector;

f) detection of the host cells which are positive (positive colonies) for the expression of the export and/or secretion marker, and/or of the promoter activity marker;

g) isolation of the DNA from the positive colonies and insertion of this DNA into a cell which is identical to that in step c);

h) selection of the inserts contained in the vector, allowing the production of clones which are positive for the export and/or secretion marker, and/or for the promoter activity marker;

i) isolation and characterization of the mycobacterial DNA fragments contained in these inserts.

In one of the preferred embodiments of the screening method according to the invention, the host cells, detected in step f), which are positive for the export and/or secretion marker are, optionally in a second stage, tested for the capacity of the selected nucleotide insert to stimulate the expression of the promoter activity marker when said host cells are phagocytosed by macrophage-type cells.

More specifically, the stimulation of the expression of the promoter activity marker in host cells placed in axenic culture (host cells alone in culture) is compared with the stimulation of the expression of the promoter activity marker in host cells cultured in the presence of macrophages and which are thus phagocytosed by the latter.

The selection of host cells which are positive for the promoter activity marker can be carried out immediately after step e) of the method of screening described above, or alternatively after any one of steps f), g), h) or i), that is to say once the host cells have been positively selected for the export and/or selection marker.

The use of this method allows the construction of DNA libraries comprising sequences corresponding to polypeptides which are capable of being exported and/or secreted, and/or which are capable of being induced or repressed during the infection when they are produced inside recombinant mycobacteria. Step i) of the method may comprise a step for sequencing the inserts selected.

Preferably, in the method according to the invention, the vector used is chosen from the plasmids pJVEDa (CNCM, No. I-1797), pJVEDb (CNCM, No. I-1906), pJVEDc (CNCM, No. I-1799) or pJVED/*M. tuberculosis* (CNCM, No. I-1907), and the digestion of the mycobacterial DNA sequences is carried out by means of the enzyme Sau3A.

According to a preferred embodiment of the invention, the method of screening is characterized in

that the mycobacterial sequences are derived from a pathogenic mycobacterium, for example from *M. tuberculosis, M. bovis, M. avium, M. africanum* or *M. leprae*.

The invention also comprises a library of genomic DNA or of cDNA which is complementary to mycobacterial mRNA, characterized in that it is obtained by a method comprising steps a) and b) or a), b) and c) of the preceding method according to the invention, preferably a library of genomic DNA or of cDNA which is complementary to mRNA of pathogenic mycobacteria, preferably of mycobacteria belonging to the *Mycobacterium tuberculosis* complex group, preferably of *Mycobacterium tuberculosis*.

In the present invention, "nucleic sequences" or "amino acid sequences" are understood to designate SEQ ID No. X to SEQ ID No. Y, where X and Y may independently represent a number or an alphanumeric character, respectively the set of nucleic sequences or the set of amino acid sequences represented by figures X to Y, ends included.

For example, the nucleic sequences or the amino acid sequences SEQ ID No. 1 to SEQ ID No. 4N are respectively the nucleic sequences or the amino acid sequences represented by figures 1 to 4N, that is to say respectively the nucleic sequences or the amino acid sequences SEQ ID No. 1, SEQ ID No. 1A', SEQ ID No. 1B', SEQ ID No. 1C', SEQ ID No. 1D, SEQ ID No. 1F, SEQ ID No. 2, SEQ ID No. 3A, SEQ ID No. 3B, SEQ ID No. 3C, SEQ ID No. 4A, SEQ ID No. 4B, SEQ ID No. 4C, SEQ ID No. 4A', SEQ ID No. 4B', SEQ ID No. 4C', SEQ ID No. 4F, SEQ ID No. 4J, SEQ ID No. 4K, SEQ ID No. 4L, SEQ ID No. 4M and SEQ ID No. 4N.

The subject of the invention is also the nucleotide sequences of mycobacteria or comprising nucleotide sequences of mycobacteria selected after carrying out the method according to the invention which is described above.

Preferably, said mycobacterium is chosen from M. tuberculosis, M. bovis, M. africanum, M. avium, M. leprae, M. paratuberculosis, M. kansassi or M. xenopi.

The nucleotide sequences of mycobacteria or comprising a mycobacterial nucleotide sequence are preferred, said mycobacterial nucleotide sequence being chosen from the sequences of mycobacterial DNA fragments having the nucleic sequence SEQ ID No. 1 to SEQ ID No. 24C, SEQ ID No. 27A to SEQ ID No. 27C, SEQ ID No. 29 and SEQ ID No. 31A to SEQ ID No. 50F, which are represented respectively by Figures 1 to 24C (plates 1 to 150), by Figures 27A to 27C (plates 152 to 154), by Figure 29 (plate 156) and by Figures 31A to 50F (plates 158 to 275).

According to a specific embodiment of the invention, preferred sequences are, for example, the mycobacterial DNA fragments having the sequence SEQ ID No. 1, SEQ ID No. 3A, SEQ ID No. 5A, SEQ ID No. 6A, SEQ ID No. 7A, SEQ ID No. 8A, SEQ ID No. 9A, SEQ ID No. 10A, SEQ ID No. 27A or SEQ ID No. 29 which are contained respectively in the vectors pDP428 (CNCM, No. I-1818), p6D7 (CNCM, No. I-1814), p5F6 (CNCM, No. I-1816), p2A29 (CNCM, No. I-1817), p5B5 (CNCM, No. I-1819), p1C7 (CNCM, No. I-1820), p2D7 (CNCM, No. I-1821), p1B7 (CNCM, No. I-1843), p5A3 (CNCM, No. I-1815) and pM1C25 (CNCM, No. I-2062).

The invention also relates to a nucleic acid comprising the entire open reading frame of one of the nucleotide sequences according to the invention, in particular one of the sequences SEQ ID No. 1 to SEQ ID No. 24C, SEQ ID No. 27A, SEQ ID No. 27C, SEQ ID No. 29 and SEQ ID No. 31A to SEQ ID No. 50F according to the invention. Said nucleic acid may be isolated, for example, in the following manner:

a) preparation of a cosmid library from the *M. tuberculosis* DNA, for example according to the technique described by Jacobs et al., 1991;

b) hybridization of all or part of a probe nucleic acid having the sequence chosen, for example, from SEQ ID No. 1 to SEQ ID No. 24C, SEQ ID No. 27A to SEQ ID No. 27C, SEQ ID No. 29 and SEQ ID No. 31A to SEQ ID No. 50F with the cosmids of the library previously prepared in step a);

c) selection of the cosmids hybridizing with the probe nucleic acid of step b);

d) sequencing of the DNA inserts of the clones selected in step c) and identification of the complete open reading frame;

e) where appropriate, cloning of the inserts sequenced in step d) into an appropriate expression and/or cloning vector.

The nucleic acids comprising the entire open reading frame of the sequences SEQ ID No. 1 to SEQ ID No. 24C, SEQ ID No. 27A to SEQ ID No. 27C, SEQ ID No. 29 and SEQ ID No. 31A to SEQ ID No. 50F are among the preferred nucleic acids.

The present invention makes it possible to determine a gene fragment encoding an exported polypeptide. Comparison with the genome sequence published by Cole et al. (Cole et al., 1998, Nature, 393, 537-544) makes it possible to determine the whole gene carrying the identified sequence according to the present invention.

Nucleotide sequence comprising the entire open reading frame of a sequence according to the invention is understood to mean the nucleotide sequence (genomic, cDNA, semisynthetic or synthetic) comprising one of the sequences according to the invention and extending, on the one hand, in 5′ of these sequences up to the first codon for initiation of translation (ATG or GTG) or even up to the first stop codon, and, on the other hand, in 3′ of these sequences up to the next stop codon, this being in any one of the three possible reading frames.

The nucleotide sequences which are complementary to the above sequences according to the invention also form part of the invention.

Polynucleotide having a sequence which is complementary to a nucleotide sequence according to the invention is understood to mean any DNA or RNA sequence whose nucleotides are complementary to those of said sequence according to the invention and whose orientation is reversed.

The nucleotide fragments of the above sequences according to the invention, which are in particular useful as probes or primers, also form part of the invention.

The invention also relates to the polynucleotides, characterized in that they comprise a polynucleotide chosen from:

a) a polynucleotide whose sequence is complementary to the sequence of a polynucleotide according to the invention,

b) a polynucleotide whose sequence comprises at least 50% identity with a polynucleotide according to the invention,

c) a polynucleotide which hybridizes, under high stringency conditions, with a polynucleotide sequence according to the invention,

d) a fragment of at least 8 consecutive nucleotides of a polynucleotide defined according to the invention.

The high stringency conditions as well as the percentage identity will be defined below in the present description.

When the coding sequence derived from the export and/or secretion marker gene is a sequence derived from the *phoA* gene, the export and/or secretion of the product of the *phoA* gene, truncated where appropriate, is obtained only when this sequence is inserted in phase with the sequence or element for regulating the expression of the production of polynucleotides and its location placed upstream, which contains the elements controlling the expression,

export and/or secretion which are derived from a mycobacterial sequence.

The recombinant vectors of the invention may of course comprise multiple cloning sites which are shifted by one or two nucleotides relative to a vector according to the invention, thus making it possible to express the polypeptide corresponding to the mycobacterial DNA fragment which is inserted and which is capable of being translated according to one of the three possible reading frames.

For example, the preferred vectors pJVEDb and pJVEDc of the invention are distinguishable from the preferred vector pJVEDa by a respective shift of one and two nucleotides at the level of the multiple cloning site.

Thus, the vectors of the invention are capable of expressing each of the polypeptides which are capable of being encoded by an inserted mycobacterial DNA fragment. Said polypeptides, characterized in that they are therefore capable of being exported and/or secreted, and/or induced or repressed, or expressed constitutively during the infection, form part of the invention.

The polypeptides of the invention whose amino acid sequences are chosen from the amino acid sequences SEQ ID No. 1 to SEQ ID No. 24C, SEQ ID No. 27A to SEQ ID No. 28 and SEQ ID No. 30 to SEQ ID No. 50F and represented respectively by Figures 1 to 24C (plates 1 to 150), Figures 27A to 28 (plates 152 to 155) and Figures 30 to 50F (plates 157 to 275) are in particular preferred.

Also forming part of the invention are the fragments or biologically active fragments as well as the polypeptides which are homologous to said polypeptides; fragment, biologically active fragment and polypeptides which are homologous to a polypeptide being as defined below in the description.

The invention also relates to the polypeptides comprising a polypeptide or one of their fragments according to the invention.

The subject of the invention is also recombinant mycobacteria containing a recombinant vector according to the invention which is described above. A preferred mycobacterium is a mycobacterium of the *M. smegmatis* type.

M. smegmatis advantageously makes it possible to test the efficiency of mycobacterial sequences for controlling the expression, export and/or secretion, and/or promoter activity of a given sequence, for example of a sequence encoding a marker such as alkaline phosphatase and/or luciferase.

Another preferred mycobacterium is a mycobacterium of the *M. bovis* type, for example the BCG strain which is currently used for vaccination against tuberculosis.

Another preferred mycobacterium is a strain of *M. tuberculosis*, *M. bovis* or *M. africanum* potentially possessing all the appropriate regulatory systems.

The inventors have thus characterized, in particular, a polynucleotide consisting of a nucleotide sequence which is present in all the tested strains of mycobacteria belonging to the *Mycobacterium tuberculosis* complex. This polynucleotide, called *DP428*, contains an open reading frame (ORF) encoding a polypeptide of about 12 kD. The open reading frame (ORF) encoding the polypeptide DP428 extends from the nucleotide at position nt 941 to the nucleotide at position nt 1351 of the sequence SEQ ID No. 2, the polypeptide DP428 having the following amino acid sequence SEQ ID No. 28:

MKTGTATTRRRLLAVLIALALPGAAVALLAEPSATGASDPCAASEVARTVGSVAK
SMGDYLDSHPETNQVMTAVLQQQVGPGSVASLKAHFEANPKVASDLHALSQPLTD
LSTRCSLPISGLQAIGLMQAVQGARR.

This molecular weight (MW) corresponds to the theoretical MW of the mature protein obtained after cleavage of the signal sequence, the MW of the protein

or polypeptide DP428 being about 10 kD after potential anchorage to peptidoglycan and potential cleavage between S and G of the LPISG motif.

This polynucleotide includes, on the one hand, an open reading frame corresponding to a structural gene and, on the other hand, the signals for regulating the expression of the coding sequence upstream and downstream of the latter. The polypeptide DP428 is composed of a signal peptide, a hydrophilic central region and a hydrophobic C-terminal region. The latter ends with two arginine residues (R), a retention signal, and is preceded by an LPISG motif which resembles the LPXTG motif for anchorage to peptidoglycan (Schneewind et al., 1995).

Structural gene for the purposes of the present invention is understood to mean a polynucleotide encoding a protein, a polypeptide or alternatively a fragment of the latter, said polynucleotide comprising only the sequence corresponding to the open reading frame (ORF), which excludes the sequences on the 5′ side of the open reading frame (ORF) which direct the initiation of transcription.

Thus, the invention relates in particular to a polynucleotide whose sequence is chosen from the nucleotide sequences SEQ ID No. 1 to SEQ ID No. 2.

More particularly, the invention relates to a polynucleotide, characterized in that it comprises a polynucleotide chosen from:

a) a polynucleotide whose sequence is chosen from the nucleotide sequences SEQ ID No. 1 to SEQ ID No. 2,

b) a polynucleotide whose nucleic sequence is the sequence between the nucleotide at position nt 964 and the nucleotide at position nt 1234, ends included, of the sequence SEQ ID No. 1,

c) a polynucleotide whose sequence is complementary to the sequence of a polynucleotide defined in a) or b),

d) a polynucleotide whose sequence exhibits at least 50% identity with a polynucleotide defined in a), b) or c),

e) a polynucleotide which hybridizes, under high stringency conditions, with a sequence of a polynucleotide defined in a), b), c) or d),

f) a fragment of at least 8 consecutive nucleotides of a polynucleotide defined in a), b), c), d) or e).

Nucleotide sequence, polynucleotide or nucleic acid is understood to mean, according to the present invention, a double-stranded DNA, a single-stranded DNA and products of transcription of said DNAs.

Percentage identity for the purpose of the present invention is understood to mean a percentage identity between the bases of two polynucleotides, this percentage being purely statistical and the differences between the two polynucleotides being distributed randomly and over their entire length.

Hybridization under high stringency conditions means that the temperature and ionic strength conditions are chosen such that they allow the hybridization between two complementary DNA fragments to be maintained.

By way of illustration, high stringency conditions of the hybridization step for the purposes of defining the polynucleotide fragments described above are advantageously the following:

the hybridization is carried out at a temperature which is preferably 65°C, in the presence of buffer marketed under the name rapid-hyb buffer by Amersham (RPN 1636) and 100 µg/ml of *E. coli* DNA.

The washing steps may, for example, be the following:

- two washes of 10 min, preferably at 65°C, in a 2 × SSC buffer and 0.1% SDS;
- two washes of 10 min, preferably at 65°C, in a 1 × SSC buffer and 0.1% SDS;
- one wash of 10 min, preferably at 65°C, in a 0.1 × SSC buffer and 0.1% SDS.

1 x SSC corresponds to 0.15 M NaCl and 0.05 M Na citrate and a 1 x Denhardt solution corresponds to 0.02% Ficoll, 0.02% of polyvinylpyrrolidone and 0.02% of bovine serum albumin.

Advantageously, a nucleotide fragment corresponding to the preceding definition will have at least 8 nucleotides, preferably at least 12 nucleotides, and still more preferably at least 20 consecutive nucleotides of the sequence from which it is derived. The high stringency hybridization conditions described above for a polynucleotide having a size of about 200 bases will be adjusted by persons skilled in the art for oligonucleotides with a larger or a smaller size, according to the teaching of Sambrook et al., 1989.

For the conditions for using the restriction enzymes with the aim of obtaining nucleotide fragments of the polynucleotides according to the invention, reference will be advantageously made to the manual by Sambrook et al., 1989.

Advantageously, a polynucleotide of the invention will contain at least one sequence comprising the stretch of nucleotides going from the nucleotide at position nt 964 to the nucleotide nt 1234 of the polynucleotide having the sequence SEQ ID No. 1.

The subject of the present invention is a polynucleotide according to the invention, characterized in that its nucleic sequence hybridizes with the DNA of a sequence of mycobacteria and preferably with the DNA of a sequence of mycobacteria belonging to the *Mycobacterium tuberculosis* complex.

The polynucleotide is encoded by a polynucleotide sequence as described supra.

The subject of the present invention is also a polypeptide derived from a mycobacterium, characterized in that it is present only in the mycobacteria belonging to the *Mycobacterium tuberculosis* complex.

The invention also relates to a polypeptide characterized in that it comprises a polypeptide chosen from:

a) a polypeptide whose amino acid sequence is included in an amino acid sequence chosen from the amino acid sequences SEQ ID No. 1 to SEQ ID No. 24C, SEQ ID No. 27A to SEQ ID No. 28 and SEQ ID No. 30 to SEQ ID No. 50F,

b) a polypeptide which is homologous to the polypeptide defined in a),

c) a fragment of at least 5 amino acids of a polypeptide defined in a) or b),

d) a biologically active fragment of a polypeptide defined in a), b) or c).

The subject of the present invention is also a polypeptide whose amino acid sequence is included in the amino acid sequence SEQ ID No. 1 or SEQ ID No. 2, or a polypeptide having the amino acid sequence SEQ ID No. 28.

Homologous polypeptide will be understood to designate the polypeptides exhibiting, relative to the natural polypeptide according to the invention such as the polypeptide DP428, certain modifications such as in particular a deletion, addition or substitution of at least one amino acid, a truncation, an extension, a chimeric fusion, and/or a mutation. Among the homologous polypeptides, those whose amino acid sequence exhibits at least 30%, preferably 50%, homology with the amino acid sequences of the polypeptides according to the invention are preferred. In the case of a substitution, one or more consecutive or nonconsecutive amino acids are replaced with "equivalent" amino acids. The expression "equivalent" amino acid is intended here to designate any amino acid capable of being substituted for one of the amino acids of the parent structure without, however, essentially modifying the immunogenic properties of the corresponding peptides. In other words, the equivalent amino acids will be those which allow the production of a

polypeptide having a modified sequence which allows the induction *in vivo* of antibodies or of cells capable of recognizing the polypeptide whose amino acid sequence is included in the amino acid sequence of the polypeptide according to the invention, such as the amino acid sequences SEQ ID No. 1 to SEQ ID No. 2, or a polypeptide having the amino acid sequence SEQ ID No. 28 (polypeptide DP428) or one of its above-defined fragments.

These equivalent aminoacyls may be determined either based on their structural homology with the aminoacyls for which they are substituted, or on the results of cross-immunogenicity assays to which the different peptides are capable of giving rise.

By way of example, there may be mentioned the possibilities of substitutions which are capable of being made without resulting in a profound modification of the immunogenicity of the corresponding modified peptides, the replacements, for example, of leucine with valine or isoleucine, of aspartic acid with glutamic acid, of glutamine with asparagine and of arginine with lysine, and the like, it being possible to naturally envisage the reverse substitutions under the same conditions.

Biologically active fragment will be understood to designate in particular a fragment of an amino acid sequence of a polypeptide having at least one of the characteristics of the polypeptides according to the invention, in particular in that it is:

- capable of being exported and/or secreted by a mycobacterium, and/or of being induced or repressed during infection with the mycobacterium; and/or
- capable of inducing, repressing or modulating, directly or indirectly, a mycobacterium virulence factor; and/or
- capable of inducing an immunogenicity reaction directed against mycobacteria; and/or
- capable of being recognized by an antibody which is specific for mycobacterium.

Polypeptide fragment is understood to designate a polypeptide comprising a minimum of 5 amino acids, preferably 10 amino acids and 15 amino acids.

A polypeptide of the invention, or one of its fragments, as defined above, is capable of being specifically recognized by the antibodies present in the serum of patients infected by mycobacteria and preferably bacteria belonging to the *Mycobacterium tuberculosis* complex or by cells of the infected host.

Thus, forming part of the invention are the fragments of the polypeptide whose amino acid sequence is included in the amino acid sequence of a polypeptide according to the invention, such as the amino acid sequences SEQ ID No. 1 to SEQ ID No. 2, or a polypeptide having an amino acid sequence SEQ ID No. 28, which may be obtained by cleavage of said polypeptide with a proteolytic enzyme, such as trypsin or chymotrypsin or collagenase, or with a chemical reagent, such as cyanogen bromide (CNBr) or alternatively by placing a polypeptide according to the invention such as the polypeptide DP428 in a very acidic environment, for example at pH 2.5. Preferred peptide fragments according to the invention, for use in diagnosis or in vaccination, are the fragments contained in regions of a polypeptide according to the invention such as the polypeptide DP428 which are capable of being naturally exposed to the solvent and to thus exhibit substantial immunogenicity properties. Such peptide fragments may be prepared either by chemical synthesis, from hosts transformed with an expression vector according to the invention containing a nucleic acid allowing the expression of said fragments, placed under the control of appropriate regulatory and/or expression elements or alternatively by chemical or enzymatic cleavage.

Analysis of the hydrophilicity of the polypeptide DP428 was carried out with the aid of the DNA Strider™ software (marketed by CEA Saclay) on the basis of a calculation of the hydrophilic character of the region encoding DP428 of SEQ ID No. 28. The results of

this analysis are presented in Figure 54 where the hydrophilicity index is detailed, for each of the amino acids (AA) having a defined position in SEQ ID No. 28. The higher the hydrophilicity index, the more the amino acid considered is likely to be exposed to the solvent in the native molecule, and is subsequently likely to exhibit a high degree of antigenicity. Thus, a stretch of at least seven amino acids possessing a high hydrophilicity index (>0.3) can constitute the basis of the structure of an immunogenic candidate peptide according to the present invention.

The cellular immune responses of the host to a polypeptide according to the invention can be demonstrated according to the techniques described by Colignon et al., 1996.

From the data of the hydrophilicity map presented in figure 54, the inventors were able to define regions of the polypeptide DP428 which are preferably exposed to the solvent, more particularly the region located between amino acids 55 and 72 of the sequence SEQ ID No. 28 and the region located between amino acids 99 and 107 of SEQ ID No. 28.

The peptide regions of the polypeptide DP428 which are defined above may be advantageously used for the production of immunogenic compositions or of vaccine compositions according to the invention.

The polynucleotides characterized in that they encode a polypeptide according to the invention also form part of the invention.

The invention also relates to the nucleic acid sequences which can be used as probes or primers, characterized in that said sequences are chosen from the nucleic acid sequences of polynucleotides according to the invention.

The invention relates, in addition, to the use of a nucleic acid sequence of polynucleotides according to the invention as a probe or a primer for the detection and/or amplification of a nucleic acid sequence. Among these nucleic acid sequences according

to the invention which can be used as probes or primers there are preferred the nucleic acid sequences of the invention, characterized in that said sequences are sequences, or their complementary sequence, between the nucleotide at position nt 964 and the nucleotide at position nt 1234, ends included, of the sequence SEQ ID No. 1.

Among the polynucleotides according to the invention which can be used as nucleotide primers, the polynucleotides having the sequences SEQ ID No. 25 and SEQ ID No. 26 are particularly preferred.

The polynucleotides according to the invention may thus be used to select nucleotide primers, in particular for the PCR technique (Erlich, 1989; Innis et al., 1990, and, Rolfs et al., 1991).

This technique requires the choice of oligonucleotide pairs flanking the fragment which has to be amplified. Reference may be made, for example, to the technique described in American patent US No. 4,683,202. These oligodeoxyribonucleotide or oligoribonucleotide primers advantageously have a length of at least 8 nucleotides, preferably of at least 12 nucleotides, and still more preferably of at least 20 nucleotides. Primers having a length of between 8 and 30 and preferably 12 and 22 nucleotides will be preferred in particular. One of the two primers is complementary to the (+) strand [forward primer] of the template and the other primer is complementary to the (-) strand [backward primer]. It is important that the primers do not possess a secondary structure or sequences which are complementary to each other. Moreover, the length and the sequence of each primer should be chosen so that the primers do not hybridize with other nucleic acids from prokaryotic or eukaryotic cells, in particular with the nucleic acids from other pathogenic mycobacteria, or with human DNA or RNA which may possibly contaminate the biological sample.

The results presented in Figure 51 show that the sequence encoding the polypeptide DP428

(SEQ ID No. 28) is not found in the DNAs of *M. fortuitum*, *M. simiae*, *M. avium*, *M. chelonae*, *M. flavescens*, *M. gordonae*, *M. marinum* and *M. kansasii*.

The amplified fragments may be identified after agarose or polyacrylamide gel electrophoresis or after capillary electrophoresis, or alternatively after a chromatographic technique (gel filtration, hydrophobic chromatography or ion-exchange chromatography). The specificity of the amplification may be checked by molecular hybridization using, as probes, the nucleotide sequences of polynucleotides of the invention, plasmids containing these sequences or their amplification products.

The amplified nucleotide fragments may be used as reagents in hybridization reactions in order to detect the presence, in a biological sample, of a target nucleic acid having a sequence which is complementary to that of said amplified nucleotide fragments.

Among the polynucleotides according to the invention which can be used as nucleotide probes, the polynucleotide fragment comprising the sequence between the nucleotide at position nt 964 and the nucleotide at position nt 1234, ends included, of the sequence SEQ ID No. 1 is most particularly preferred.

These probes and amplicons may be labeled or otherwise with radioactive elements or with non-radioactive molecules such as enzymes or fluorescent elements.

The invention also relates to the nucleotide fragments which are capable of being obtained by amplification with the aid of primers according to the invention.

Other techniques for the amplification of the target nucleic acid may be advantageously used as alternatives to PCR.

The SDA (Strand Displacement Amplification) technique (Walker et al., 1992) is an isothermic amplification technique whose principle is based on the

capacity of a restriction enzyme to cut one of the two strands of its recognition site which is in the form of a hemiphosphorothioate and on the property of a DNA polymerase to initiate the synthesis of a new DNA strand from the 3'OH end created by the restriction enzyme and to displace the strand previously synthesized which is present downstream.

The polynucleotides of the invention, in particular the primers according to the invention, may also be used in other methods of amplifying a target nucleic acid, such as:

- the TAS (Transcription-based Amplification System) technique described by Kwoh et al. in 1989;
- the 3SR (Self-Sustained Sequence Replication) technique described by Guatelli et al. in 1990;
- the NASBA (Nucleic Acid Sequence Based Amplification) technique described by Kievitis et al. in 1991;
- the TMA (Transcription Mediated Amplification) technique.

The polynucleotides of the invention may also be used in techniques for the amplification or modification of the nucleic acid serving as probe, such as:

- the LCR (Ligase Chain Reaction) technique described by Landegren et al. in 1988 and improved by Barany et al. in 1991, which uses a heat-stable ligase;
- the RCR (Repair Chain Reaction) technique described by Segev in 1992;
- the CPR (Cycling Probe Reaction) technique described by Duck et al. in 1990;
- the Q-beta-replicase amplification technique described by Miele et al. in 1983 and improved in particular by Chu et al. in 1986, Lizardi et al. in 1988 and then by Burg et al. as well as Stone et al. in 1996.

In the case where the target polynucleotide to be detected is an RNA, for example an mRNA, a reverse transcriptase-type enzyme will be advantageously used,

prior to using an amplification reaction using the primers according to the invention or to the use of a method of detection using the probes of the invention, in order to obtain a cDNA from the RNA contained in the biological sample. The cDNA obtained will then serve as target for the primers or probes used in the method of amplification or detection according to the invention.

The detection probe will be chosen so that it hybridizes with the amplicon generated. Such a detection probe will advantageously have a sequence of at least 12 nucleotides in particular of at least 15 nucleotides and preferably at least 200 nucleotides.

The nucleotide probes according to the invention are capable of detecting mycobacteria and preferably bacteria belonging to the *Mycobacterium tuberculosis* complex, more particularly because of the fact that these mycobacteria possess in their genome at least one copy of polynucleotides according to the invention. These probes according to the invention are capable, for example, of hybridizing with the nucleotide sequence of a polypeptide according to the invention, more particularly any oligonucleotide hybridizing with the sequence SEQ ID No. 1 encoding the *M. tuberculosis* polypeptide DP428 and not exhibiting a cross-hybridization reaction or an amplification reaction (PCR) with, for example, sequences present in mycobacteria not belonging to the *Mycobacterium tuberculosis* complex. The nucleotide probes according to the invention hybridize specifically with a DNA or RNA molecule of a polynucleotide according to the invention, under high stringency hybridization conditions as given in the form of an example above.

The nonlabeled sequences may be used directly as probes. However, the sequences are generally labeled with a radioactive element (^{32}P, ^{35}S, ^{3}H, ^{125}I) or with a nonradioactive molecule (biotin, acetylaminofluorene, digoxigenin, 5-bromodeoxyuridine, fluorescein) in order to obtain probes which can be used for many applications.

Examples of nonradioactive labelings of probes are described, for example, in French patent No. 78,10975 or by Urdea et al. or by Sanchez-Pescador et al. in 1988.

In the latter case, it will also be possible to use one of the labeling methods described in patents FR 2,422,956 and FR 2,518,755. The hybridization technique may be carried out in various ways (Matthews et al., 1988). The most common method consists in immobilizing the nucleic acid extracted from mycobacterial cells onto a support (such as nitrocellulose, nylon, polystyrene) and in incubating, under well-defined conditions, the immobilized target nucleic acid with the probe. After hybridization, the excess probe is removed and the hybrid molecules formed are detected by the appropriate method (measurement of the radioactivity, of the fluorescence or of the enzymatic activity linked to the probe).

Advantageously, the labeled nucleotide probes according to the invention may have a structure such that they make amplification of the radioactive or nonradioactive signal possible. An amplification system corresponding to the above definition will comprise detection probes in the form of a branched, ramified DNA such as those described by Urdea et al. in 1991. According to this technique, several types of probe, in particular a capture probe, to immobilize the target DNA or RNA to a support, and a detection probe will be advantageously used. The detection probe binds a "branched" DNA having a ramified structure. The branched DNA in turn is capable of binding oligonucleotide probes which are themselves coupled to alkaline phosphatase molecules. The activity of this enzyme is then detected using a chemiluminescent substrate, for example a derivative of dioxethane phosphate.

According to another advantageous embodiment of the nucleic probes according to the invention, they can be covalently or noncovalently immobilized on a support

and used as capture probes. In this case, a probe termed "capture probe" is immobilized on a support and serves to capture, through specific hybridization, the target nucleic acid obtained from the biological sample to be tested. If necessary, the solid support is separated from the sample and the duplex formed between the capture probe and the target nucleic acid is then detected by means of a second probe termed "detection probe" which is labeled with an easily detectable element.

The oligonucleotide fragments may be obtained from the sequences according to the invention by cleavage with restriction enzymes or by chemical synthesis according to conventional methods, for example according to the method described in European patent No. EP-0,305,929 (Millipore Corporation) or by other methods.

An appropriate method of preparing the nucleic acids of the invention comprising a maximum of 200 nucleotides (or 200 bp in the case of double-stranded nucleic acids) comprises the following steps:

- synthesis of DNA using the automated beta-cyanethylphosphoramidite method described in 1986,
- cloning of the nucleic acids thus obtained into an appropriate vector and recovery of the nucleic acids by hybridization with an appropriate probe.

A method of preparation, by the chemical route, of nucleic acids according to the invention having a length greater than 200 nucleotides (or 200 bp in the case of double-stranded nucleic acids) comprises the following steps:

- assembly of chemically synthesized oligonucleotides, provided at their end with different restriction sites, whose sequences are compatible with the stretch of amino acids of the natural polypeptide according to the principle described in 1983,
- cloning of the nucleic acids thus obtained into an appropriate vector and recovery of the desired nucleic acids by hybridization with an appropriate probe.

The nucleotide probes used for recovering the desired nucleic acids in the abovementioned methods generally consist of 8 to 200 nucleotides of the polypeptide sequence according to the invention and are capable of hybridizing with the nucleic acid tested for under the hybridization conditions defined above. The synthesis of these probes may be carried out according to the automated beta-cyanethylphosphoramidite method described in 1986.

The oligonucleotide probes according to the invention may be used in a detection device comprising an oligonucleotide array library. An exemplary embodiment of such an array library may consist of an array of probe oligonucleotides which are attached to a support, the sequence of each probe of a given length being situated with a shift of one or more bases relative to the preceding probe, each of the probes of the array arrangement thus being complementary to a distinct sequence of the target DNA or RNA to be detected and each probe of known sequence being attached at a predetermined position of the support. The target sequence to be detected may be advantageously labeled radioactively or nonradioactively. When the labeled target sequence is brought into contact with the array device, it forms hybrids with the probes having complementary sequences. A nuclease treatment, followed by washing, makes it possible to remove the probe-target sequence hybrids which are not perfectly complementary. Because of the precise knowledge of the sequence of a probe at a given position of the array, it is then possible to deduce the nucleotide sequence of the target DNA or RNA sequence. This technique is particularly effective when matrices of oligonucleotide probes of a large size are used.

An alternative to the use of a labeled target sequence may consist of using a support allowing a "bioelectronic" detection of the hybridization of the target sequence with the probes of the array support,

when said support consists of or comprises a material capable of acting, for example, as an electron donor at the positions of the array where a hybrid has been formed. Such an electron-donating material is for example gold. The detection of the nucleotide sequence of the target DNA or RNA is then determined by an electronic device.

An exemplary embodiment of a biosensor, as defined above, is described in European patent application No. EP-0,721,016 in the name of Affymax Technologies N.V. or in American patent No. US 5,202,231 in the name of Drmanac.

The subject of the invention is also the hybrid polynucleotides resulting:

- either from the formation of a hybrid molecule between an RNA or a DNA (genomic DNA or cDNA) obtained from a biological sample with a probe or a primer according to the invention,
- or from the formation of a hybrid molecule between an RNA or a DNA (genomic DNA or cDNA) obtained from a biological sample with a nucleotide fragment amplified with the aid of a pair of primers according to the invention.

cDNA for the purposes of the invention is understood to mean a DNA molecule obtained by causing a reverse transcriptase type enzyme to act on an RNA molecule, in particular a messenger RNA (mRNA) molecule, according to the techniques described in Sambrook et al. in 1989.

The subject of the present invention is also a family of recombinant plasmids, characterized in that they contain at least one nucleotide sequence of a polynucleotide according to the invention. According to an advantageous embodiment of said plasmid it comprises the nucleotide sequence SEQ ID No. 1 or a fragment thereof.

Another subject of the present invention is a vector for the cloning, expression and/or insertion of a sequence, characterized in that it comprises a

nucleotide sequence of a polynucleotide according to the invention at a site which is not essential for its replication, where appropriate under the control of regulatory elements capable of playing a role in the expression of the polypeptide DP428, in a given host.

Specific vectors are for example plasmids, phages, cosmids, phagemids and YACs.

These vectors are useful for transforming host cells so as to clone or express the nucleotide sequences of the invention.

The invention also comprises the host cells transformed with a vector according to the invention.

Preferably, the host cells are transformed under conditions allowing the expression of a recombinant polypeptide according to the invention.

A preferred host cells according to the invention is the *E. coli* strain transformed with the plasmid pDP428 deposited on 28 January 1997 at the CNCM under the No. I-1818 or transformed with the plasmid pM1C25 which was deposited on 4 August 1998 at the CNCM under the No. I-2062 or a mycobacterium belonging to a strain of *M. tuberculosis*, *M. bovis* or *M. africanum* potentially possessing all the appropriate regulatory systems.

It is now easy to produce proteins or polypeptides in a relatively large quantity by genetic engineering using, as expression vectors, plasmids, phages or phagemids. All or part of the *DP428* gene, or any polynucleotide according to the invention, may be inserted into an appropriate expression vector in order to produce *in vitro* a polypeptide according to the invention, in particular the polypeptide DP428. Said polypeptide may be attached to a microplate in order to develop a serological test intended to search, for diagnostic purposes, for the specific antibodies in patients suffering tuberculosis.

Thus, the present invention relates to a method of preparing a polypeptide, characterized in that it uses a vector according to the invention. More

particularly, the invention relates to a method of preparing a polypeptide of the invention comprising the following steps:

- where appropriate, the prior amplification, according to the PCR technique, of the quantity of nucleotide sequences encoding said polypeptide with the aid of two DNA primers chosen so that one of these primers is identical to the first 10 to 25 nucleotides of the nucleotide sequence encoding said polypeptide, while the other primer is complementary to the last 10 to 25 nucleotides (or hybridizes with these last 10 to 25 nucleotides) of said nucleotide sequence, or conversely so that one of these primers is identical to the last 10 to 25 nucleotides of said sequence, while the other primer is complementary to the first 10 to 25 nucleotides (or hybridizes with the first 10 to 25 nucleotides) of said nucleotide sequence, followed by the introduction of said sequences thus amplified into an appropriate vector,
- the culture, in an appropriate culture medium, of a cellular host which has been previously transformed with an appropriate vector containing a nucleic acid according to the invention comprising the nucleotide sequence encoding said polypeptide, and
- the separation, from said culture medium, of said polypeptide produced by said transformed cellular host.

The subject of the invention is also a polypeptide which is capable of being obtained by a method of the invention as described above.

The peptides according to the invention may also be prepared by techniques which are conventionally used in the field of peptide synthesis. This synthesis may be carried out in homogeneous solution or in solid phase.

For example, the technique of synthesis in homogeneous solution described by Houbenweyl in 1974 will be used.

This method of synthesis consists in successively condensing in pairs the successive aminoacyls in

the required order, or in condensing aminoacyls and fragments formed beforehand and already containing several aminoacyls in the appropriate order, or alternatively several fragments thus prepared beforehand, it being understood that care will be taken to protect beforehand all the reactive functions carried by these aminoacyls or fragments, with the exception of the amine functions of one and the carboxyl functions of the other or vice versa, which should normally be involved in the formation of the peptide bonds, in particular after activation of the carboxyl function, according to methods well known in peptide synthesis. As a variant, use may be made of coupling reactions using conventional coupling reagents, of the carbodiimide type, such as for example 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide.

When the aminoacyl used possesses an additional acid function (in particular in the case of glutamic acid), these functions will be protected, for example with t-butyl ester groups.

In the case of gradual synthesis, amino acid by amino acid, the synthesis preferably starts with the condensation of the C-terminal amino acid with the amino acid which corresponds to the neighboring aminoacyl in the desired sequence, and so on, step by step, up to the N-terminal amino acid.

According to another preferred technique of the invention, the one described by Merrifield is used.

To manufacture a peptide chain according to the Merrifield method, use is made of a very porous polymer resin onto which the first C-terminal amino acid of the chain is attached. This amino acid is attached to the resin via its carboxyl group and its amine function is protected, for example with the t-butyloxycarbonyl group.

When the first C-terminal amino acid is thus attached to the resin, the group for protecting the amine function is removed by washing the resin with an acid.

In the case where the group for protecting the amine function is the t-butyloxycarbonyl group, it may be removed by treating the resin with trifluoroacetic acid.

The second amino acid which provides the second aminoacyl of the desired sequence, from the C-terminal aminoacyl residue, is then coupled with the deprotected amine function of the first C-terminal amino acid attached to the chain. Preferably, the carboxyl function of this second amino acid is activated, for example with dicyclohexylcarbodiimide, and the amine function is protected, for example with t-butyloxycarbonyl.

The first portion of the desired peptide chain is thus obtained which comprises two amino acids, and whose terminal amine function is protected. As before, the amine function is deprotected and it is then possible to proceed to the attachment of the third aminoacyl, under conditions similar to those for the addition of the second C-terminal amino acid.

The amino acids which will constitute the peptide chain will thus be attached, one after the other, to the amino group, each time deprotected beforehand, of the portion of the peptide chain which is already formed and which is attached to the resin.

When the entire desired peptide chain is formed, the groups for protecting the different amino acids constituting the peptide chain are removed and the peptide is detached from the resin, for example with the aid of hydrofluoric acid.

Preferably, said polypeptides which are capable of being obtained by a method of the invention as described above will comprise a region exposed to the solvent and will have a length of at least 20 amino acids.

According to another embodiment of the invention, said polypeptides are specific to mycobacteria of the *Mycobacterium tuberculosis* complex and are not

therefore recognized by antibodies specific for other mycobacterial proteins.

The invention relates, in addition, to hybrid polypeptides having at least one polypeptide according to the invention and a sequence of a polypeptide capable of inducing an immune response in humans or animals.

Advantageously, the antigenic determinant is such that it is capable of inducing a humoral and/or cellular response.

Such a determinant may comprise a polypeptide according to the invention, in glycosylated form, which is used to obtain immunogenic compositions capable of inducing the synthesis of antibodies directed against multiple epitopes. Said glycosylated polypeptides also form part of the invention.

These hybrid molecules may consist in part of a polypeptide-carrying molecule according to the invention combined with a portion, in particular an epitope of the diphtheria toxin, the tetanus toxin, a hepatitis B virus surface antigen (patent FR 79 21811), the VP1 antigen of the poliomyelitis virus or any other viral or bacterial toxin or antigen.

Advantageously, said antigenic determinant corresponds to an antigenic determinant of immunogenic proteins of 45/47 kD of *M. tuberculosis* (international application PCT/FR 96/0166), or alternatively which are selected for example from ESAT6 (Harboe et al., 1996, Andersen et al., 1995, and Sorensen et al., 1995) and DES (PCT/FR 97/00923, Gicquel et al.).

A viral antigen, as defined above, will be preferably a hepatitis virus surface or envelope protein, for example the hepatitis B surface protein in one of its S,S-preS1, S-preS2 or S-preS2-preS1 forms or alternatively a protein of a hepatitis A virus, or of a hepatitis non-A, non-B virus, such as a hepatitis C, E or delta virus.

More particularly, a viral antigen as defined above will be the whole or part of one of the glyco-

proteins encoded by the genome of the HIV-1 virus (patents GB 8324800, EP 84401834 or EP 85905513) or of the HIV-2 virus (EP 87400151), and in particular the whole or part of a protein selected from gag, pol, nef or env of HIV-1 or HIV-2.

The methods for synthesizing the hybrid molecules include the methods used in genetic engineering to construct hybrid polynucleotides encoding the desired polypeptide sequences. Reference may be advantageously made, for example, to the technique for the production of genes encoding fusion proteins described by Minton in 1984.

Said hybrid polynucleotides encoding a hybrid polypeptide as well as the hybrid polypeptides according to the invention characterized in that they are recombinant proteins obtained by the expression of said hybrid polynucleotides also form part of the invention.

The polypeptides according to the invention may advantageously be used in a method for the *in vitro* detection of antibodies directed against said polypeptides, in particular the polypeptide DP428, and also of antibodies directed against a bacterium of the *Mycobacterium tuberculosis* complex, in a biological sample (biological tissue or fluid) capable of containing them, this method comprising bringing this biological sample into contact with a polypeptide according to the invention under conditions allowing an immunological reaction *in vitro* between said polypeptide and the antibodies which may be present in the biological sample, and detecting *in vitro* the antigen-antibody complexes which may be formed.

The polypeptides according to the invention may also and advantageously be used in a method for the detection of an infection by a bacterium of the *Mycobacterium tuberculosis* complex in a mammal based on the *in vitro* detection of a cellular reaction indicating prior sensitization of the mammal to said polypeptide such as for example cell proliferation, the synthesis of proteins such as interferon-gamma. This

method for the detection of an infection by a bacterium of the *Mycobacterium tuberculosis* complex in a mammal is characterized in that it comprises the following steps:

a) preparation of a biological sample containing cells of said mammal, more particularly cells of the immune system of said mammal and still more particularly T cells;

b) incubation of the biological sample of step a) with a polypeptide according to the invention;

c) detection of a cellular reaction indicating prior sensitization of the mammal to said polypeptide such as for example cell proliferation and/or the synthesis of proteins such as interferon-gamma.

Cell proliferation may be measured, for example, by incorporation of ^{3}H-Thymidine.

Also forming part of the invention are the methods for the detection of a delayed hypersensitivity reaction (DTH), characterized in that they use a polypeptide according to the invention.

Preferably, the biological sample consists of a fluid, for example a human or animal serum, blood, biopsies, bronchoalveolar fluid or pleural fluid.

Any conventional procedure may be used to carry out such a detection.

By way of example, a preferred method uses immunoenzymatic procedures such as the ELISA, immunofluorescence or radioimmunoassay (RIA) technique and the like.

Thus, the invention also relates to the polypeptides according to the invention, labeled with the aid of a suitable marker of the enzymatic, fluorescent or radioactive type.

Such methods comprise, for example, the following steps:

- deposition of predetermined quantities of a polypeptide composition according to the invention into the wells of a microtiter plate,

- introduction into said wells of increasing dilutions of serum or of another biological sample as defined above, before being analyzed,
- incubation of the microplate,
- introduction into the wells of the microtiter plate of labeled antibodies directed against human or animal immunoglobulins, the labeling of these antibodies having been carried out with the aid of an enzyme selected from those which are capable of hydrolyzing a substrate while modifying its radiation absorption, at least at a defined wavelength, for example at 550 nm,
- detection, by comparing with a control, of the quantity of substrate hydrolyzed.

The invention also relates to a box or kit for the *in vitro* diagnosis of an infection by a mycobacterium belonging to the *Mycobacterium tuberculosis* complex, comprising:

- a polypeptide according to the invention,
- where appropriate, the reagents for constituting the medium which is appropriate for the immunological or specific reaction,
- the reagents allowing the detection of the antigen-antibody complexes produced by the immunological reaction which may be present in the biological sample, and the *in vitro* detection of the antigen-antibody complexes which may be formed, it being possible for these reagents to also carry a marker, or to be capable of being recognized in turn by a labeled reagent, more particularly in the case where the polypeptide according to the invention is not labeled,
- where appropriate, a reference biological sample (negative control) free of antibodies recognized by a polypeptide according to the invention,
- where appropriate, a reference biological sample (positive control) containing a predetermined quantity of antibodies recognized by a polypeptide according to the invention.

The polypeptides according to the invention make it possible to prepare monoclonal or polyclonal antibodies which are characterized in that they recognize specifically the polypeptides according to the invention. The monoclonal antibodies may be advantageously prepared from hybridomas according to the technique described by Kohler and Milstein in 1975. The polyclonal antibodies may be prepared, for example, by immunizing an animal, in particular a mouse, with a polypeptide according to the invention combined with an immune response adjuvant, and then purifying the specific antibodies contained in the serum of the immunized animals on an affinity column to which the polypeptide which served as antigen has been attached beforehand. The polyclonal antibodies according to the invention may also be prepared by purifying an affinity column, to which there have been immobilized beforehand a polypeptide according to the invention, antibodies contained in the serum of patients infected with a mycobacterium and preferably a bacterium belonging to the *Mycobacterium tuberculosis* complex.

The subject of the invention is also mono- or polyclonal antibodies or fragments thereof, or chimeric antibodies, characterized in that they are capable of recognizing specifically a polypeptide according to the invention.

The antibodies of the invention may also be labeled in the same manner as described above for the nucleic probes of the invention, such as a labeling of the enzymatic, fluorescent or radioactive type.

The invention relates, in addition, to a method for the specific detection of the presence of an antigen of a mycobacterium and preferably a bacterium of the *Mycobacterium tuberculosis* complex in a biological sample, characterized in that it comprises the following steps:

a) bringing the biological sample (biological tissue or fluid) collected from an individual into contact with a mono- or polyclonal antibody according to the

invention, under conditions allowing an immunological reaction *in vitro* between said antibodies and the polypeptides specific to mycobacteria and preferably bacteria of the *Mycobacterium tuberculosis* complex which may be present in the biological sample, and

b) detection of the antigen-antibody complex formed.

Also coming within the scope of the invention is a box or kit for the *in vitro* diagnosis, on a biological sample, of the presence of strains of mycobacteria and preferably of bacteria belonging to the *Mycobacterium tuberculosis* complex, preferably *M. tuberculosis*, characterized in that it comprises:

- a polyclonal or monoclonal antibody according to the invention, labeled where appropriate;
- where appropriate, a reagent for constituting the medium which is appropriate for carrying out the immunological reaction;
- a reagent allowing the detection of the antigen-antibody complexes produced by the immunological reaction, it being possible for this reagent to also carry a marker, or to be capable of being recognized in turn by a labeled reagent, more particularly in the case where said monoclonal or polyclonal antibody is not labeled;
- where appropriate, reagents for carrying out the lysis of the cells of the sample tested.

The subject of the present invention is also a method for the detection and rapid identification of the mycobacteria and preferably of the *M. tuberculosis* bacteria in a biological sample, characterized in that it comprises the following steps:

a) isolation of the DNA from the biological sample to be analyzed, or production of a cDNA from the RNA of the biological sample;

b) specific amplification of the DNA of mycobacteria and preferably of bacteria belonging to the *Mycobacterium tuberculosis* complex with the aid of primers according to the invention;

c) analysis of the products of amplification.

The products of amplification may be analyzed by various methods.

Two methods of analysis are given by way of example below:

- agarose gel electrophoretic analysis of the products of amplification. The presence of a DNA fragment which migrates to the expected position suggests that the sample analyzed contained DNA of mycobacteria belonging to the tuberculosis complex, or
- analysis by the molecular hybridization technique using a nucleic probe according to the invention. This probe will be advantageously labeled with a nonradioactive (cold probe) or radioactive element.

For the purposes of the present invention, "DNA of the biological sample" or "DNA contained in the biological sample" is understood to mean either the DNA present in the biological sample considered, or the cDNA obtained after the action of a reverse transcriptase-type enzyme on the RNA present in said biological sample.

Another method of the present invention allows the detection of an infection by a mycobacterium and preferably a bacterium of the *Mycobacterium tuberculosis* complex in a mammal. This method comprises the following steps:

a) preparation of a biological sample containing cells of said mammal, more particularly cells of the immune system of said mammal and still more particularly T cells;

b) incubation of the biological sample of step a) with a polypeptide according to the invention;

c) detection of a cellular reaction indicating prior sensitization of the mammal to said polypeptide in particular cell proliferation and/or the synthesis of proteins such as interferon-gamma;

d) detection of a reaction of delayed hypersensitivity or of sensitization of the mammal to said polypeptide.

This method of detection is an intradermal method which is described for example by M.J. Elhay et al. (1988) Infection and Immunity, 66(7): 3454-3456.

Another aim of the present invention consists in a method for the detection of the mycobacteria and preferably the bacteria belonging to the *Mycobacterium tuberculosis* complex in a biological sample, characterized in that it comprises the following steps:

a) bringing an oligonucleotide probe according to the invention into contact with a biological sample, the DNA contained in the biological sample, or the cDNA obtained by reverse transcription of the RNA of the biological sample, having, where appropriate, been made accessible to the hybridization beforehand, under conditions allowing the hybridization of the probe with the DNA or the cDNA of the mycobacteria and preferably of the bacteria of the *Mycobacterium tuberculosis* complex;

b) detection of the hybrid formed between the oligonucleotide probe and the DNA of the biological sample.

The invention also relates to a method for the detection of the mycobacteria and preferably of the bacteria belonging to the *Mycobacterium tuberculosis* complex in a biological sample, characterized in that it comprises the following steps:

a) bringing an oligonucleotide probe according to the invention, immobilized on a support, into contact with a biological sample, the DNA of the biological sample having, where appropriate, been made accessible to the hybridization beforehand, under conditions allowing the hybridization of said probe with the DNA of the mycobacteria and preferably of the bacteria of the *Mycobacterium tuberculosis* complex;

b) bringing the hybrid formed between said oligonucleotide probe immobilized on a support and the DNA contained in the biological sample, where appropriate after removal of the DNA of the biological sample which

has not hybridized with the probe, into contact with a labeled oligonucleotide probe according to the invention.

According to an advantageous embodiment of the method of detection defined above, it is characterized in that, prior to step a), the DNA of the biological sample is amplified beforehand with the aid of a pair of primers according to the invention.

Another embodiment of the method of detection according to the invention consists in a method for the detection of the presence of the mycobacteria and preferably the bacteria belonging to the *Mycobacterium tuberculosis* complex in a biological sample, characterized in that it comprises the following steps:

a) bringing the biological sample into contact with a pair of primers according to the invention, the DNA contained in the sample having been, where appropriate, made accessible to hybridization beforehand, under conditions allowing hybridization of said primers with the DNA of the mycobacteria and preferably of the bacteria of the *Mycobacterium tuberculosis* complex;

b) amplification of the DNA of a mycobacterium and preferably of a bacterium of the *Mycobacterium tuberculosis* complex;

c) detection of the amplification of the DNA fragments corresponding to the fragment flanked by the primers, for example by gel electrophoresis or by means of an oligonucleotide probe according to the invention.

A subject of the invention is also a method for the detection of the presence of the mycobacteria and preferably the bacteria belonging to the *Mycobacterium tuberculosis* complex in a biological sample by strand displacement, characterized in that it comprises the following steps:

a) bringing the biological sample into contact with two pairs of primers according to the invention specifically intended for amplification of the SDA type described above, the DNA content in the sample having been, where appropriate, made accessible to hybridiza-

tion beforehand, under conditions allowing hybridization of the primers with the DNA of the mycobacteria and preferably the bacteria of the *Mycobacterium tuberculosis* complex;

b) amplification of the DNA of the mycobacteria and preferably of the bacteria of the *Mycobacterium tuberculosis* complex;

c) detection of the amplification of DNA fragments corresponding to the fragment flanked by the primers, for example by gel electrophoresis or by means of an oligonucleotide probe according to the invention.

The invention also relates to a box or kit for carrying out the method described above, intended for the detection of the presence of the mycobacteria and preferably the bacteria of the *Mycobacterium tuberculosis* complex in a biological sample, characterized in that it comprises the following components:

a) an oligonucleotide probe according to the invention;

b) the reagents necessary for carrying out a hybridization reaction;

c) where appropriate, a pair of primers according to the invention as well as the reagents necessary for a reaction of amplification of the DNA (genomic DNA, plasmid DNA or cDNA) of mycobacteria and preferably of bacteria of the *Mycobacterium tuberculosis* complex.

The subject of the invention is also a kit or box for the detection of the presence of the mycobacteria and preferably the bacteria of the *Mycobacterium tuberculosis* complex in a biological sample, characterized in that it comprises the following components:

a) an oligonucleotide probe, termed capture probe, according to the invention;

b) an oligonucleotide probe, termed revealing probe, according to the invention;

c) where appropriate, a pair of primers according to the invention as well as the reagents necessary for a

reaction of amplification of the DNA of mycobacteria and preferably of bacteria of the *Mycobacterium tuberculosis* complex.

The invention also relates to a kit or box for the amplification of the DNA of the mycobacteria and preferably the bacteria of the *Mycobacterium tuberculosis* complex present in a biological sample, characterized in that it comprises the following components:

a) a pair of primers according to the invention;

b) the reagents necessary for carrying out a DNA amplification reaction;

c) optionally, a component which makes it possible to verify the sequence of the amplified fragment, more particularly an oligonucleotide probe according to the invention.

Another subject of the present invention relates to an immunogenic composition, characterized in that it comprises a polypeptide according to the invention.

Another immunogenic composition according to the invention is characterized in that it comprises one or more polypeptides according to the invention and/or one or more hybrid polypeptides according to the invention.

According to an advantageous embodiment, the above-defined immunogenic composition constitutes a vaccine when it is provided in combination with a pharmaceutically acceptable vehicle and optionally one or more immunity adjuvants such as alum or a representative of the family of muramyl peptides or alternatively incomplete Freund's adjuvant.

Various types of vaccine are currently available for protecting humans against infectious diseases: attenuated live microorganisms (*M. bovis* - BCG for tuberculosis), inactivated microorganisms (influenza virus), acellular extracts (*Bordetella pertussis* for whooping cough), recombinant proteins (hepatitis B virus surface antigen), polysaccharides

(pneumococci). Experiments are being carried out on vaccines prepared from synthetic peptides or genetically modified microorganisms expressing heterologous antigens. More recently still, recombinant plasmid DNAs carrying genes encoding protective antigens have been proposed as an alternative vaccine strategy. This type of vaccination is carried out with a specific plasmid which is derived from an *E. coli* plasmid which does not replicate *in vivo* and which encodes only the vaccinal protein. The principal functional components of this plasmid are: a strong promoter allowing expression in eukaryotic cells (for example that of CMV), an appropriate cloning site for inserting the gene of interest, a termination-polyadenylation sequence, a prokaryotic replication origin for producing the recombinant plasmid *in vitro* and a selectable marker (for example the ampicillin-resistance gene) for facilitating the selection of the bacteria which contain the plasmid. Animals were immunized by simply injecting the naked plasmid DNA into the muscle. This technique leads to the expression of the vaccinal protein *in situ* and to an immune response in particular of the cellular type (CTL) and of the humoral type (antibody). This double induction of the immune response is one of the main advantages of the vaccination technique with naked DNA. Huygen et al. (1996) and Tascon et al. (1996) succeeded in obtaining a degree of protection against *M. tuberculosis* by injecting recombinant plasmids containing *M. leprae* genes (*hsp65, 36kDa pra*) as inserts. *M. leprae* is the agent responsible for leprosy. The use of an insert specific to *M. tuberculosis* such as, for example, the whole or part of the *DP428* gene, which is the subject of the present invention, would probably lead to a better protection against tuberculosis. The whole or part of the *DP428* gene, or any polynucleotide according to the invention, can be easily inserted into the plasmid vectors V1J (Montgomery et al., 1993), pcDNA3 (Invitrogen, R & D Systems) or pcDNA1/Neo (Invitrogen)

which possess the necessary characteristics for a vaccinal use.

The invention thus relates to a vaccine, characterized in that it comprises one or more polypeptides according to the invention and/or one or more hybrid polypeptides according to the invention as previously defined, in combination with a pharmaceutically compatible vehicle and, where appropriate, one or more appropriate immunity adjuvants.

The invention also relates to a vaccine composition intended for the immunization of humans or animals against a bacterial or viral infection, such as tuberculosis or hepatitis, characterized in that it comprises one or more hybrid polypeptides as previously defined in combination with a pharmaceutically compatible vehicle and, where appropriate, one or more immunity adjuvants.

Advantageously, in the case of a protein which is a hybrid between a polypeptide according to the invention and the hepatitis B surface antigen, the vaccine composition will be administered, in humans, in an amount of 0.1 to 1 µg of purified hybrid protein per kilogram of the weight of the patient, preferably 0.2 to 0.5 µg/kg of the weight of the patient, for a dose intended for a given administration. In the case of patients suffering from disorders of the immune system, in particular immunosuppressed patients, each injected dose will preferably contain half of the quantity, by weight, of the hybrid protein contained in a dose intended for a patient not suffering from immune system disorders.

Preferably, the vaccine composition will be administered several times, spread out over time, by the intradermal or subcutaneous route. By way of example, three doses as defined above will be administered, respectively, to the patient at time t0, at time t0 + 1 month and at time t0 + 1 year.

Alternatively, three doses will be administered, respectively, to the patient at time t0, at time t0 + 1 month and at time t0 + 6 months.

In mice, in which a weight dose of the vaccine composition comparable to the dose used in humans is administered, the antibody reaction is tested by collecting serum followed by a study of the formation of a complex between the antibodies present in the serum and the antigen of the vaccine composition, according to the customary techniques.

The invention also relates to an immunogenic composition characterized in that it comprises a polynucleotide or an expression vector according to the invention, in combination with a vehicle allowing its administration to humans or animals.

The subject of the invention is also a vaccine intended for immunizing against a bacterial or viral infection, such as tuberculosis or hepatitis, characterized in that it comprises a polynucleotide or an expression vector according to the invention, in combination with a pharmaceutically acceptable vehicle.

Such immunogenic or vaccine compositions are in particular described in international application No. WO 90/11092 (Vical Inc.) and also in international application No. WO 95/11307 (Institut Pasteur).

The constituent polynucleotide of the immunogenic composition or of the vaccine composition according to the invention may be injected into the host after having been coupled with compounds which promote the penetration of this polynucleotide into the cell or its transport to the cell nucleus. The resulting conjugates may be encapsulated into polymer microparticles, as described in international application No. WO 94/27238 (Medisorb Technologies International).

According to another embodiment of the immunogenic and/or vaccine composition according to the invention, the polynucleotide, preferably a DNA, is complexed with DEAE-dextran (Pagano et al., 1967) or with nuclear proteins (Kaneda et al., 1989), with

lipids (Felgner et al., (1987) or encapsulated into liposomes (Fraley et al., 1980).

According to yet another advantageous embodiment of the immunogenic and/or vaccine composition according to the invention, the polynucleotide according to the invention may be introduced in the form of a gel facilitating its transfection into cells. Such a composition in gel form may be a poly-L-lysine and lactose complex, as described by Midoux in 1993, or Poloxamer 407™, as described by Pastore in 1994. The polynucleotide or the vector according to the invention may also be in suspension in a buffer solution or may be combined with liposomes.

Advantageously, such a vaccine will be prepared in accordance with the technique described by Tacson et al. or Huygen et al. in 1996 or in accordance with the technique described by Davis et al. in international application No. WO 95/11307 (Whalen et al.).

Such a vaccine will be advantageously prepared in the form of a composition containing a vector according to the invention, placed under the control of regulatory elements allowing its expression in humans or animals.

To produce such a vaccine, the polynucleotide according to the invention is first of all subcloned into an appropriate expression vector, particularly an expression vector containing regulatory and expression signals recognized by the enzymes in eukaryotic cells and also containing a replication origin which is active in prokaryotes, for example in *E. coli*, which allows its prior amplification. The purified recombinant plasmid obtained is then injected into the host, for example by the intramuscular route.

It will be possible, for example, to use as vector for expressing *in vivo* the antigen of interest the plasmid pcDNA3 or the plasmid pcDNA1/neo, both marketed by Invitrogen (R&D Systems, Abingdon, United Kingdom). It is also possible to use the plasmid V1Jns.tPA described by Shiver et al. in 1995.

Such a vaccine will advantageously comprise, in addition to the recombinant vector, a saline solution, for example a sodium chloride solution.

A vaccine composition as defined above will be, for example, administered by the parenteral route or by the intramuscular route.

The present invention also relates to a vaccine characterized in that it contains one or more nucleotide sequences according to the invention and/or one or more polynucleotides as mentioned above in combination with a pharmaceutically compatible vehicle and, where appropriate, one or more appropriate immunity adjuvants.

Another aspect relates to a method of screening molecules capable of inhibiting the growth of mycobacteria or the maintenance of mycobacteria in a host, characterized in that said molecules block the synthesis or the function of the polypeptides encoded by a nucleotide sequence according to the invention or by a polynucleotide as described supra.

In said method of screening, the molecules may be anti-messengers or may induce the synthesis of anti-messengers.

The present invention also relates to molecules capable of inhibiting the growth of mycobacteria or the maintenance of mycobacteria in a host, characterized in that said molecules are synthesized based on the structure of the polypeptides encoded by a nucleotide sequence according to the invention or by a polynucleotide as described supra.

Other characteristics and advantages of the invention appear in the following examples and figures:

FIGURES

The Figure 1 series:

The Figure 1 series illustrates the series of nucleotide sequences SEQ ID No. 1 corresponding to the insert

of the vector pDP428 (deposited at the CNCM under the No. I-1818) and the series of amino acid sequences SEQ ID No. 1 of the polypeptides encoded by the series of nucleotide sequences SEQ ID No. 1.

Figure 2:

Illustrates the nucleotide sequence SEQ ID No. 2 corresponding to the region including the gene encoding the polypeptide DP428 (region underlined). Both the ATG and GTG codons for initiation of translation were taken into account in this figure. The figure shows that the polypeptide DP428 is probably part of an operon comprising at least three genes. The double-boxed region probably includes the promoter regions.
The single-boxed region corresponds to the motif LPISG which resembles the motif LPXTG described in Gram-positive bacteria as allowing anchorage to peptidoglycans.

The Figure 3 series:

The Figure 3 series represents the series of nucleotide sequences SEQ ID No. 3 corresponding to the insert of the vector p6D7 (deposited at the CNCM under the No. I-1814).

The Figure 4 series:

The Figure 4 series represents the series of nucleotide sequences SEQ ID No. 4 corresponding to the insert of the vector p5A3 (deposited at the CNCM under the No. I-1815).

The Figure 5 series:

The Figure 5 series represents the series of nucleotide sequences SEQ ID No. 5 corresponding to the insert of

the vector p5F6 (deposited at the CNCM under the No. I-1816).

The Figure 6 series:

The Figure 6 series represents the series of nucleotide sequences SEQ ID No. 6 corresponding to the insert of the vector p2A29 (deposited at the CNCM under the No. I-1817).

The Figure 7 series:

The Figure 7 series represents the series of nucleotide sequences SEQ ID No. 7 corresponding to the insert of the vector p5B5 (deposited at the CNCM under the No. I-1819).

The Figure 8 series:

The Figure 8 series represents the series of nucleotide sequences SEQ ID No. 8 corresponding to the insert of the vector p1C7 (deposited at the CNCM under the No. I-1820).

The Figure 9 series:

The Figure 9 series represents the series of nucleotide sequences SEQ ID No. 9 corresponding to the insert of the vector p2D7 (deposited at the CNCM under the No. I-1821).

The Figure 10 series:

The Figure 10 series represents the series of nucleotide sequences SEQ ID No. 10 corresponding to the insert of the vector p1B7 (deposited at the CNCM under the No. I-1843).

The Figure 11 series:

The Figure 11 series represents the series of nucleotide sequences SEQ ID No. 11.

The Figure 12 series:

The Figure 12 series represents the series of nucleotide sequences SEQ ID No. 12.

The Figure 13 series:

The Figure 13 series represents the series of nucleotide sequences SEQ ID No. 13.

The Figure 14 series:

The Figure 14 series represents the series of nucleotide sequences SEQ ID No. 14 corresponding to the insert of the vector p5B5 (deposited at the CNCM under the No. I-1819).

The Figure 15 series:

The Figure 15 series represents the series of nucleotide sequences SEQ ID No. 15.

The Figure 16 series:

The Figure 16 series represents the series of nucleotide sequences SEQ ID No. 16.

The Figure 17 series:

The Figure 17 series represents the series of nucleotide sequences SEQ ID No. 17.

The Figure 18 series:

The Figure 18 series represents the series of nucleotide sequences SEQ ID No. 18.

The Figure 19 series:

The Figure 19 series represents the series of nucleotide sequences SEQ ID No. 19.

The Figure 20 series:

The Figure 20 series represents the series of nucleotide sequences SEQ ID No. 20 corresponding to the insert of the vector p2A29 (deposited at the CNCM under the No. I-1817).

The Figure 21 series:

The Figure 21 series represents the series of nucleotide sequences SEQ ID No. 21.

The Figure 22 series:

The Figure 22 series represents the series of nucleotide sequences SEQ ID No. 22.

The Figure 23 series:

The Figure 23 series represents the series of nucleotide sequences SEQ ID No. 23.

The Figure 24 series:

The Figure 24 series represents the series of nucleotide sequences SEQ ID No. 24.

Figures 25 and 26:

Figures 25 and 26 illustrate, respectively, the sequences SEQ ID No. 25 and SEQ ID No. 26 representing a pair of primers used to specifically amplify, by PCR, the region corresponding to nucleotides 964 to 1234 included in the sequence SEQ ID No. 1.

The Figure 27 series:

The Figure 27 series represents the series of nucleotide sequences SEQ ID No. 27 corresponding to the insert of the vector p5A3.

Figure 28:

The amino acid sequence as defined in Figure 28 represents the amino acid sequence SEQ ID No. 28 corresponding to the polypeptide DP428.

Figure 29:

Figure 29 represents the nucleotide sequence SEQ ID No. 29 of the complete gene encoding the M1C25 protein.

Figure 30:

Figure 30 represents the amino acid sequence SEQ ID No. 30 of the M1C25 protein.

The Figure 31 series:

The Figure 31 series represents the series of nucleotide sequences SEQ ID No. 31.

The Figure 32 series:

The Figure 32 series represents the series of nucleotide sequences SEQ ID No. 32.

The Figure 33 series:

The Figure 33 series represents the series of nucleotide sequences SEQ ID No. 33.

The Figure 34 series:

The Figure 34 series represents the series of nucleotide sequences SEQ ID No. 34.

The Figure 35 series:

The Figure 35 series represents the series of nucleotide sequences SEQ ID No. 35.

The Figure 36 series:

The Figure 36 series represents the series of nucleotide sequences SEQ ID No. 36.

The Figure 37 series:

The Figure 37 series represents the series of nucleotide sequences SEQ ID No. 37.

The Figure 38 series:

The Figure 38 series represents the series of nucleotide sequences SEQ ID No. 38.

The Figure 39 series:

The Figure 39 series represents the series of nucleotide sequences SEQ ID No. 39.

The Figure 40 series:

The Figure 40 series represents the series of nucleotide sequences SEQ ID No. 40.

The Figure 41 series:

The Figure 41 series represents the series of nucleotide sequences SEQ ID No. 41 corresponding to the insert of the vector p2D7 (deposited at the CNCM under the No. I-1821).

The Figure 42 series:

The Figure 42 series represents the series of nucleotide sequences SEQ ID No. 42.

The Figure 43 series:

The Figure 43 series represents the series of nucleotide sequences SEQ ID No. 43.

The Figure 44 series:

The Figure 44 series represents the series of nucleotide sequences SEQ ID No. 44.

The Figure 45 series:

The Figure 45 series represents the series of nucleotide sequences SEQ ID No. 45.

The Figure 46 series:

The Figure 46 series represents the series of nucleotide sequences SEQ ID No. 46.

The Figure 47 series:

The Figure 47 series represents the series of nucleotide sequences SEQ ID No. 47.

The Figure 48 series:

The Figure 48 series represents the series of nucleotide sequences SEQ ID No. 48.

The Figure 49 series:

The Figure 49 series represents the series of nucleotide sequences SEQ ID No. 49.

The Figure 50 series:

The Figure 50 series represents the series of nucleotide sequences SEQ ID No. 50.

Figure 51:

A. Construct pJVED: shuttle plasmid (capable of multiplying in mycobacteria as well as in *E. coli*) with a kanamycin-resistance gene (derived from Tn*903*) as a selectable marker. The truncated *phoA* gene (Δ *phoA*) and the *luc* gene form a synthetic operon.
B. Joining sequence between *phoA* and *luc*.

Figure 52:

Genomic hybridization (Southern blotting) of the genomic DNA of various mycobacterial species with the aid of an oligonucleotide probe whose sequence is the sequence between the nucleotide at position nt 964 (5′ end of the probe) and the nucleotide at position nt 1234 (3′ end of the probe), ends included, of the sequence SEQ ID No. 1.

Figures 53 and 54:

Recombinant *M. smegmatis* Luc and PhoA activities containing pJVED with various nucleotide fragments as described in the examples. Figures 52 and 53 represent the results obtained for two separate experiments carried out under the same conditions.

Figure 55:

Representation of the hydrophobicity (Kyte and Doolitle) of the coding sequence of the polypeptide DP428 with its schematic representation. The LPISG motif immediately precedes the hydrophobic C-terminal region. The sequence ends with two arginines.

Figure 56:

Representation of the hydrophobicity (Kyte and Doolitle) of the sequence of the polypeptide M1C25 having the amino acid sequence SEQ ID No. 30.

Figure 57:

A - Acrylamide gel (12%) under denaturing conditions of a bacterial extract obtained by sonication of *E. coli* M15 bacteria containing the plasmid pM1C25 without and after 4 hours of induction with IPTG, stained with Comassie blue.

Lane 1: Molecular weight marker (Prestained SDS-PAGE Standards High Range BIO-RAD®).

Lane 2: Bacterial extract obtained by sonication of *E. coli* M15 bacteria containing the plasmid pM1C25 without induction with IPTG.

Lane 3: Bacterial extract obtained by sonication of *E. coli* M15 bacteria containing the plasmid pM1C25 after 4 hours of induction with IPTG.

Lane 4: Molecular weight marker (Prestained SDS-PAGE Standards Low Range BIO-RAD®).

B - Western blotting of a gel similar gel (12% acrylamide) visualized by means of the penta-His antibody marketed by the company Quiagen.

Lane 1: Representation of the molecular weight marker (Prestained SDS-PAGE Standards High Range BIO-RAD®).

Lane 2: Bacterial extract obtained by sonication of *E. coli* M15 bacteria containing the plasmid pM1C25 without induction with IPTG.

Lane 3: Bacterial extract obtained by sonication of *E. coli* M15 bacteria containing the plasmid pM1C25 after 4 hours of induction with IPTG.

Lane 4: Representation of the molecular weight marker (Prestained SDS-PAGE Standards Low Range BIO-RAD®).

The band which is most predominantly present in the lanes corresponding to the bacteria induced with IPTG compared with those not induced with IPTG, between 34,200 and 28,400 daltons, corresponds to the expression of the insert M1C25 cloned into the vector pQE-60 (Qiagen®).

As regards the legend to the other figures which are numbered by an alphanumeric character, each of these other figures represents the nucleotide sequence and the amino acid sequence having the SEQ ID sequence whose numbering is identical to the alphanumeric character of each of said figures.

The alphanumeric numberings of the figures representing the SEQ IDs comprising a number followed by a letter have the following meanings:

- the alphanumeric numberings having the same number relate to the same family of sequences attached to the reference SEQ ID sequence whose numbering has this same number and the letter A;
- the letters A, B and C for the same family of sequences distinguish the three possible reading frames of the reference SEQ ID nucleotide sequence (A);
- the letters with a prime (') index mean that the sequence corresponds to a fragment of the reference SEQ ID sequence (A);
- the letter D means that the sequence corresponds to the sequence of the gene predicted by Cole et al., 1998;
- the letter F means that the sequence corresponds to the open reading frame (ORF) containing the corresponding "D" sequence according to Cole et al., 1998;
- the letter G means that the sequence is a sequence predicted by Cole et al., 1998, and exhibiting a homology of more than 70% with the reference SEQ ID sequence (A);
- the letter H means that the sequence corresponds to the open reading frame containing the corresponding "G" sequence according to Cole et al., 1998;
- the letter R means that the sequence corresponds to a sequence predicted by Cole et al., 1998, upstream of the corresponding "D" sequence and capable of being in phase with the sequence "D" because of possible sequencing errors;
- the letter P means that the sequence corresponds to the open reading phase containing the corresponding "R" sequence;
- the letter Q means that the sequence corresponds to a sequence containing the corresponding "F" and "P" sequences.

As regards the sequence family SEQ ID No. 4, the preceding insert *phoA* contains two fragments which are noncontiguous on the genome, SEQ ID 4J and SEQ ID 4A, and which are therefore derived from a multiple cloning allowing the expression and export of *phoA*. These two noncontiguous fragments, the genes and the open reading frames containing them according to Cole et al., 1998, are important for the export of an antigenic polypeptide:

- the letters J, K and L distinguish the three possible reading frames of the corresponding nucleotide sequence "J";
- the letter M means that the sequence corresponds to the sequence predicted by Cole et al., 1998, and containing the sequence SEQ ID No. 4J;
- the letter N means that the sequence corresponds to the open reading frame containing the sequence SEQ ID No. 4M.

As regards the sequence family SEQ ID No. 45, the letter Z means that the sequence corresponds to the sequence of a cloned fragment fused with *phoA*.

Finally, as regards the sequence family SEQ ID No. 41, the letter S means that the sequence corresponds to a sequence predicted by Cole et al., 1998 and which may be in the same reading frame as the corresponding sequence "D", the letter T meaning that the corresponding sequence contains the corresponding sequences "F" and "S".

EXAMPLES

Materials and methods

Bacterial cultures, plasmids and culture media

E. coli was cultured on Luria-Bertani (LB) solid or liquid medium. *M. smegmatis* was cultured on Middlebrook 7H9 liquid medium (Difco) supplemented with albumin-dextrose (ADC), 0.2% glycerol and 0.05% Tween,

or on solid L medium. If necessary, the antibiotic kanamycin was added at a concentration of 20 µg/ml. The bacterial clones having a PhoA activity were detected on LB agar containing 5-bromo-4-chloro-3-indolyl phosphate (X-P, at 40 µg/ml).

Manipulation of DNA and sequencing

The manipulations of DNA and the Southern-blot analyses were carried out using the standard techniques (Sambrook et al., 1989). The double-stranded DNA sequences were determined with a Taq Dye Deoxy Terminator Cycle sequencing kit (Applied Biosystems), in a System 9600 GeneAmp PCR (Perkin-Elmer), and after migration on a model 373 DNA analyzing system (Applied Biosystems).

Constructions of the plasmids

The plasmid pJVED$_a$ was constructed from pLA71, a transfer plasmid comprising the *phoA* gene which is truncated and placed in phase with *BlaF*. pLA71 was cleaved with the restriction enzymes *Kpn*I and *Not*I, thus removing *phoA* without affecting the promoter of *BlaF*. The *luc* gene encoding the firefly luciferase was amplified from pGEM-*luc* and a ribosome-binding site was added. *phoA* was amplified from pJEM11. The amplified fragments were cleaved with *Pst*I and ligated together. The oligodeoxynucleotides used are the following:

pPV.luc.Fw: 5'GACTGCTGCAGAAGGAGAAGATCCAAATGG3'
luc.Bw: 5'GACTAGCGGCCGCGAATTCGTCGACCTCCGAGG3'
pJEM.phoA.Fw: 5'CCGCGGATCCGGATACGTAC3'
phoA.Bw: 5'GACTGCTGCAGTTTATTTCAGCCCCAGAGCG3'.

The fragment thus obtained was reamplified using the oligonucleotides complementary to its ends, cleaved with *Kpn*I and *Not*I, and integrated into pLA71 cleaved with the same enzymes. The resulting construct was electroporated into *E. coli* DH5α and *M. smegmatis* mc^2 155. An *M. smegmatis* clone emitting light and

having a *phoA* activity was selected and called pJVED/*blaF*. The insert was removed using *Bam*HI and the construct closed again on itself, thus reconstructing $pJVED_a$. To obtain $pJVED_{b,c}$, the multiple cloning site was cleaved with *Sca*I and *Kpn*I and closed again, removing one ($pJVED_b$) or two ($pJVED_c$) nucleotides from the *Sna*BI site. After fusion, it was thus possible to obtain six reading frames. The insert of pJVED/*hsp18* was obtained by polymerase chain amplification (PCR) of pPM1745 (Servant et al., 1995) using oligonucleotides having the sequence:

18.Fw: 5'GTACCAGTACTGATCACCCGTCTCCCGCAC3'

18.Back: AGTCAGGTACCTCGCGGAAGGGGTCAGTGCG3'

The product was cleaved with *Kpn*I and *Sca*I, and ligated to $pJVED_a$, cleaved with the same enzymes, thus leaving pJVED/*hsp18*.

pJVED/*P19kDa* and pJVED/*erp* were constructed by cleaving with *Bam*HI the insert of pExp410 and pExp53, respectively, and inserting them into the *Bam*HI site of the multiple cloning site of $pJVED_a$.

Measurement of the alkaline phosphatase activity

The presence of activity is detected by the blue color of the colonies growing on a culture medium containing the substrate 5-bromo-4-chloro-3-indolyl phosphate (XP), and then the activity can be quantitatively measured more precisely in the following manner:

M. smegmatis was cultured in an LB medium supplemented with 0.05% Tween 80 (Aldrich) and kanamycin (20 µg/ml) at 37°C for 24 hours. The alkaline phosphatase activity was measured by the Brockman and Heppel method (Brockman et al., 1968) in a sonicated extract, with *p*-nitrophenyl phosphate as reaction substrate. The quantity of proteins was measured by the Bio-Rad assay. The alkaline phosphatase activity is expressed as arbitrary units (optic density at 420 nm × µg of protein^{-1} × minutes^{-1}).

Measurement of the luciferase activity

M. smegmatis was cultured in an LB medium supplemented with 0.05% Tween 80 (Aldrich) and kanamycin (20 µg/ml) at 37°C for 24 hours and used in full exponential growth (OD at 600 nm between 0.3 and 0.8). The aliquots of bacterial suspensions were briefly sonicated and the cell extract was used to measure the luciferase activity. 25 µl of the sonicated extract were mixed with 100 µl of substrate (Promega luciferase assay system) automatically in a luminometer and the emitted light expressed in RLU (Relative Light Units). The bacteria were counted by serial dilutions of the origin suspension on LB-kanamycin agar medium and the luciferase activity expressed in RLU/µg of bacterial protein or in RLU/10^3 bacteria.

Construction of *M. tuberculosis* and *M. bovis-BCG* genomic libraries

The libraries were obtained essentially using $pJVED_{a,b,c}$, which are described above.

Preparation of macrophages derived from bone marrow and infection with recombinant *M. smegmatis*

The macrophages derived from bone marrow were prepared as described by Lang et al., 1991. In summary, the bone marrow cells were removed from the femur of 6- to 12-week old C57BL/6 mice (Iffa-Credo, France). The cells in suspensions were washed and resuspended in DMEM enriched with 10% fetal calf serum, 10% of conditioned L-cell medium and 2 mM glutamine, without antibiotics. 10^6 cells were inoculated on flat-bottomed 24-well Costar plates in 1 ml. After four days at 37°C in a humid atmosphere containing a CO_2 content of 10%, the macrophages were rinsed and reincubated for an additional two to four days. The cells of a control well were lysed with triton x 100 at 0.1% in water and

the nuclei enumerated. About 5×10^5 adherent cells were counted. For the infection, *M. smegmatis* carrying the different plasmids was cultured in full exponential phase ($OD_{6000\ nm}$ between 0.4 and 0.8) and diluted to an OD of 0.1 and then 10-fold in a medium for macrophage. 1 ml was added to each well and the plates were centrifuged and incubated for four hours at 37°C. After three washes, the cells were incubated in a medium containing amikacin for two hours. After three new washes, the adherent infected cells were incubated in a macrophage medium overnight. The cells were then lysed in 0.5 ml of lysis buffer (Promega). 100 µl were sonicated and the light emitted was measured on 25 µm. Simultaneously, the bacteria were enumerated by spreading on L-agar-kanamycin (20 µg/ml). The light emitted is expressed in RLU/10^3 bacteria.

Analyses of the databanks

The nucleotide sequences were compared with EMBL and GenBank using the FASTA algorithm and the protein sequences were analyzed by similitude by means of the PIR and Swiss Prot databanks using the BLAST algorithm.

Example 1: The pJVED vectors

The pJVED vectors (Figure 51) are plasmids carrying an *E. coli* truncated *phoA* gene without initiation codon, signal sequence and regulatory sequence. The multiple cloning site (MCS) allows the insertion of fragments of the genes encoding potential exported proteins as well as their regulatory sequences. Consequently, the fusion protein may be produced and may exhibit an alkaline phosphatase activity if it is exported. Only the fusions in phase may be produced. Thus, the MCS was modified so that the fusions may be obtained in six reading frames. The firefly luciferase *luc* gene was inserted downstream of

phoA. The complete gene with the initiation codon, but without any promoter having been used, thus ought to be expressed with *phoA* as in a synthetic operon. A new ribosome-binding site was inserted eight nucleotides upstream of the *luc* initiation codon. Two transcriptional terminators are present in the pJVED vectors, one upstream of the MCS and a second downstream of *luc*. These vectors are *E. coli*-mycobacterium transfer plasmids with a kanamycin-resistance gene as selectable marker.

phoA and luc function as in an operon, but export is necessary for the phoA activity.

Four plasmids were constructed by insertion into the MCS of DNA fragments of diverse origin:

In the first construct called pJVED/*blaF*, the 1.4 kb fragment is derived from the plasmid already described pLA71 (Lim et al., 1995). This fragment, derived from the β-lactamase gene (*blaF*) of *M. fortuitum* D216 (Timm et al., 1994), includes the hyperactive mutated promoter, the segment encoding 32 amino acids of the signal sequence and the first 5 amino acids of the mature protein. Thus, this construct includes the strongest promoter known in mycobacterium and the elements necessary for the export of the *phoA* fusion protein. Consequently, a strong light emission and a good *phoA* activity can be expected from this construct (cf. Figures 53 and 54).

Into a second construct called pJVED/*hsp18*, a 1.5 kb fragment was cloned from the plasmid already described pPM1745 (Servant et al., 1995). This fragment includes the nucleotides encoding the first ten amino acids of the 18 kb heat shock protein derived from *Streptomyces albus* (heat shock protein 18, HSP 18), the ribosome-binding site, the promoter and, upstream, regulatory sites controlling its expression. This protein belongs to the alpha-crystalline family of low-molecular weight HSP (Verbon et al., 1992). Its homolog, derived from *M. leprae*, the 18 kDa antigen, is

already known to be induced during phagocytosis by a murine macrophage of the J-774 cell line (Dellagostinet et al., 1995). Under standard culture conditions, pJVED/*hsp18* shows a weak *luc* activity and no *phoA* activity (cf. Figures 53 and 54).

In a third construct, called pJVED/*P19kDa*, the insert derived from pExp410 (Lim et al., 1995) was cleaved and cloned into the MCS of $pJVED_a$. This fragment includes the nucleotides encoding the first 134 amino acids of the *M. tuberculosis* 19 kDa known protein and of its regulatory sequences. As has been demonstrated, this protein is a glycosylated lipoprotein (Garbe et al., 1993; Herrmann et al., 1996). In Figures 53 and 54, a good *luc* activity corresponding to a strong promoter is observed for this construct, but the *phoA* activity is the strongest of the four constructs. The high *phoA* activity of this fusion protein with a lipoprotein is explained by the fact that it remains attached to the cell wall by its N-terminal end.

In the fourth and last construct, called pJVED/*erp*, the insert is derived from pExp53 (Lim et al., 1995) and was cloned into the MCS of $pJVED_a$. pExp53 is the initial plasmid selected for its *phoA* activity and containing a portion of the *M. tuberculosis erp* gene which encodes a 28-kDa antigen. The latter includes the signal sequence, a portion of the mature protein and, upstream of the initiation codon, the ribosome-binding site. The promoter was mapped. A putative iron box of the *fur* type is present in this region and flanks the -35 region of the promoter (Berthet et al., 1995). As expected (Figures 53 and 54) this construct exhibits a good light emission and a good *phoA* activity. The fact that this fusion protein, unlike the fusion with the lipoprotein of 19 kDa, does not appear to be attached to the cell wall does not exclude that the native protein is combined with it. Furthermore, the

C-terminal end of *erp* is absent from the fusion protein.

Example 2: Construction of an *M. tuberculosis* genomic DNA library in the pJVED$_s$ vectors and identification of one of the members of these libraries, (DP428), induced during phagocytosis by murine macrophages derived from bone marrow

The various constructs were tested for their capacity to evaluate the intracellular expression of the genes identified by the expression of *phoA*. For this purpose, the *luc* activity is expressed in RLU for 10^3 bacteria in axenic culture and/or under intracellular conditions. The induction or the repression following phagocytosis by the bone marrow-derived murine macrophages can be suitably evaluated by the measurement of specific activities. The results of two separate experiments are presented in Table 2.

The plasmid pJVED/*hsp18* was used as positive control for the induction during the intracellular growth phase. Although the induction of the promoter by heating the bacterium at 42°C was not conclusive, the phagocytosis of the bacterium clearly leads to an increase in the activity of the promoter. In all the experiments, the intracellular *luc* activity was strongly induced, increasing by 20 to 100-fold the initially weak basal activity (Servant, 1995).

The plasmid pJVED/*blaF* was used as a control for nonspecific modulation during the phagocytosis. It was possible to detect weak variations which were probably due to changes in culture conditions. Whatever the case, these weak variations are not comparable to the induction observed with the plasmid pJVED/*hsp18*.

All the members of the DNA library were tested by measuring the activity of the promoter during the intracellular growth. Among these, DP428 is strongly induced during phagocytosis (Tables 1 and 2).

TABLE 1

Construct	% Recovery	RLU/10^3 extracellular bacteria	RLU/10^3 intracellular bacteria	Induction
pJVED/*blaF**	0.5	1460	1727	1.2
pJVED/*hsp18*	0.6	8	57	7.1
pJVED/*DP428*	0.7	0.06	18	300

Construct	% Recovery C57BL/6	Balb/C	RLU/10^3 extracellular bacteria	RLU/10^3 intracellular bacteria C57BL/6	Balb/C	Induction C57BL/6	Balb/C
pJVED/*blaF**	7	1.1	662	250	911	0.4	1.4
pJVED/*hsp18*	6.7	1.7	164	261	325	1.6	2
pJVED/*DP428*	1.6	2.1	0.08	1.25	3.3	15.6	41

TABLE 2

Construct	% Recovery	RLU/10^3 extracellular bacteria	RLU/10^3 intracellular bacteria	Induction
pJVED/*blaF**	22	1477	367	0.25
pJVED/*hsp18*	7	0.26	6.8	26
PJVED/*DP428*	21	0.14	4	28

The nucleotide fragment encoding the N-terminal region of the polypeptide DP428 having the sequence SEQ ID No. 28 is contained in the plasmid deposited at the CNCM under the No. I-1818.

The entire sequence encoding the polypeptide DP428 was obtained as detailed below.

A probe was obtained by PCR with the aid of oligonucleotides having the sequence SEQ ID No. 25 and SEQ ID No. 26. This probe was labeled by random extension in the presence of [^{32}P]dCTP. Hybridization of the genomic DNA of *M. tuberculosis* strain Mt103 previously digested with the endonuclease ScaI was

carried out with the aid of said probe. The results of the hybridization revealed that a DNA fragment of about 1.7 kb was labeled. Because an ScaI site exists, extending from the nucleotide nt 984 to the nucleotide nt 989 of the sequence SEQ ID No. 1, that is to say on the 5′ side of the sequence used as probe, the end of the coding sequence is necessarily present in the fragment detected by hybridization.

The genomic DNA of the *M. tuberculosis* Mt 103 strain, after digestion with ScaI, was subjected to migration on agarose gel. The fragments of between 1.6 and 1.8 kb in size were cloned into the vector pSL1180 (Pharmacia) previously cleaved with ScaI and dephosphorylated. After transformation of *E. coli* with the resulting recombinant vectors, the colonies obtained were screened with the aid of the probe. The screening made it possible to isolate six colonies hybridizing with this probe.

The inserts contained in the plasmids of the previously selected recombinant clones were sequenced and then the sequences aligned so as to determine the entire sequence encoding DP428, more specifically SEQ ID No. 2.

A pair of primers were synthesized in order to amplify, starting with the genomic DNA of *M. tuberculosis*, strain Mt 103, the entire sequence encoding the polypeptide DP428. The amplicon obtained was cloned into an expression vector.

Pairs of primers appropriate for the amplification and the cloning of the sequence encoding the polypeptide DP428 can be easily produced by persons skilled in the art, on the basis of the nucleotide sequences SEQ ID No. 1 and SEQ ID No. 2.

A specific pair of primers according to the invention is the following pair of primers, which is capable of amplifying the DNA encoding the polypeptide DP428 lacking its signal sequence:

- forward primer (SEQ ID No. 29), comprising the sequence going from the nucleotide at position nt 1021 to the nucleotide nt 1044 of the sequence SEQ ID No. 2:
5′ -AGTGCATGCTGCTGGCCGAACCATCAGCGAC- 3′

- backward primer (SEQ ID No. 30), comprising the sequence complementary to the forward sequence of the nucleotide at position nt 1345 to the nucleotide at position nt 1325 of the sequence SEQ ID No. 2:
5′ -CAGCCAGATCTGCGGGCGCCCTGCACCGCCTG- 3′,

in which the portion underlined represents the sequences hybridizing specifically with the sequence SEQ ID No. 2 and the 5′ ends correspond to restriction sites for the cloning of the resulting amplicon into a cloning and/or expression vector.

A specific vector used for the expression of the polypeptide DP428 is the vector pQE70 marketed by the company Qiagen.

Example 3: The complete sequence of the *DP428* gene and its flanking regions

A probe of the coding region of DP428 was obtained by PCR and used to hybridize the genomic DNA of various mycobacterial species. According to the results of Figure 3, the gene is present only in mycobacteria of the *M. tuberculosis* complex.

Analysis of the sequence suggests that DP428 could be part of an operon. The coding sequence and the flanking regions exhibit no homology with known sequences deposited in databanks.

Based on the coding sequence, the gene encodes a 10 kDa protein with a signal peptide, a hydrophobic C-terminal end which ends with two arginines and is preceded by an LPISG motif similar to the known LPXTG motif. These two arginines could correspond to a retention signal and the protein DP428 could be attached via this motif to peptidoglycans as has

already been described in other Gram$^+$ bacteria (Navarre et al., 1994 and 1996).

The mechanism for survival and intracellular growth of mycobacteria is complex and the intimate relationships between the bacteria and the host cell remain unexplained. Whatever the mechanism, the growth and the intracellular survival of mycobacteria depend on factors produced by the bacteria produced by the bacterium and capable of modulating the response of the host. These factors may be molecules which are exposed at the cell surface, such as LAM or cell surface-associated proteins, or actively secreted molecules.

On the other hand, intracellularly, the bacteria themselves have to confront a hostile environment. They appear to respond to this by means similar to those used under stress conditions, by inducing heat shock proteins (Dellagostin et al., 1995), but also by the induction or the repression of various proteins (Lee et al., 1995). Using a methodology derived from PCR, Plum and Clark-curtiss (Plum et al., 1994) have shown that an *M. avium* gene included in a 3 kb DNA fragment is induced after phagocytosis by human macrophages. This gene encodes an exported protein comprising a leader sequence but exhibiting no significant homology with the sequences proposed by databanks. The induction, during the intracellular growth phase, of a low-molecular-weight heat shock protein derived from *M. leprae* has also been demonstrated (Dellagostin et al., 1995). In another study, the bacterial proteins from *M. tuberculosis* were metabolically labeled during the intracellular growth phase or under stress conditions and separated by two-dimensional gel electrophoresis: 16 *M. tuberculosis* proteins were induced and 28 were repressed. The same proteins are involved during stress caused by a low pH, a heat shock, H_2O_2, or during phagocytosis by human monocytes of the THP1 line. Whatever the case, the behavior of the induced and repressed proteins was unique under each condition (Lee et al., 1995). Taken

together, these results indicate that a subtle molecular dialogue is installed between the bacteria and their host cells. This dialogue probably depends on the fate of the intracellular organism.

In this context, the induction of the expression of DP428 could be of major importance, indicating an important role for this protein in intracellular survival and growth.

The method used in these experiments to evaluate the intracellular expression of the genes (cf. Jacobs et al., 1993, for the method for determining the expression of firefly luciferase, and Lim et al., 1995, for the method for determining the expression of the *PhoA* gene) has the advantage of being simple compared with the other techniques such as the technique described by Mahan et al. (Mahan et al., 1993) adapted to mycobacteria and proposed by Bange et al., (Bange et al., 1996) or the subtractive method based on PCR described by Plum and Clark-curtiss (Plum et al., 1994). Variability undoubtedly exists as shown by comparing the various experiments. Although causing the induction or the repression is sufficient, it is now possible to evaluate it, thus providing an additional tool for the physiological studies of the exported proteins identified by fusion with *phoA*.

Example 4:

Search for modulation of the activity of the promoters during the intramacrophage phases

Mouse bone marrow macrophages are prepared as described by Lang and Antoine (Lang et al., 1991). Recombinant *M. smegmatis* bacteria, whose luciferase activity per 10^3 bacteria has been determined as above, are incubated at 37°C under a humid atmosphere enriched with 5% CO_2, for 4 hours in the presence of these macrophages such that they are phagocytosed. After rinsing in order to remove the remaining extracellular bacteria, amikacin (100 µg/ml) is added to the culture

medium for two hours. After another rinsing, the medium is replaced with an antibiotic-free culture medium (DMEM enriched with 10% calf serum and 2 mM glutamine). After overnight incubation as above, the macrophages are lysed at low temperature (4°C) with the aid of a lysis buffer (cee lysis buffer, Promega), and the luciferase activity per 10^3 bacteria is determined. The ratio of the activities at placing in culture and after one night gives the coefficient of induction.

Example 5:

Isolation of a series of sequences by sequencing directly using colonies

A series of sequences allowing the expression and export of *phoA* were isolated from the DNA of *M. tuberculosis* or of *M. bovis* BCG. Among this group of sequences, two of them were further studied, the entire genes corresponding to the inserts were cloned, sequenced and antibodies against the product of these genes served to show by electron microscopy that these genes encoded antigens found at the surface of the tubercle bacilli. One of these genes, *erp*, encoding a consensus export signal sequence, the other, *des*, possessed no characteristic of a gene encoding an exported protein, based on the sequence. Another gene, DP428, was sequenced before the sequence of the *M. tuberculosis* genome became available. It contains a sequence resembling the consensus sequence for attachment to peptidoglycan, which suggests that it is also an antigen which is probably found at the surface of the tubercle bacilli. The study of the three genes, *erp*, *des*, and that encoding DP428, shows that the *phoA* system which we have developed in mycobacteria makes it possible to pick out genes encoding exported proteins with no determinant which can be picked out by studies *in silico*. This is particularly true for the polypeptides which do not possess a consensus signal

sequence (*des*) or no similarity with proteins having a known function (*erp* and *DP428*).

A number of inserts were identified and sequenced before knowing the genome of *M. tuberculosis* or of others below. These sequences may be considered as primers which make it possible to search for genes encoding exported proteins. To date, a series of primers have been sequenced and the entire corresponding genes have been either sequenced or identified based on the published sequence of the genome. To take into account sequencing errors which are always possible, the regions upstream or downstream of some primers were considered as being capable of forming part of sequences encoding exported proteins. In some cases, similarities with genes encoding exported proteins or sequences characteristic of export signals or topological characteristics of membrane proteins were detected.

Primer sequences are found to correspond to genes belonging to families of genes possessing more than 50% similarity. It is thus possible to indicate that the other genes detected by similarity with a primer encode exported proteins. This is the case for the sequence SEQ ID No. 8G and SEQ ID No. 8H which possess more than 77% similarity with SEQ ID No. 8A'.

The sequences which may encode exported proteins are the following: SEQ ID No. 1, 8, 9, 8G, 8H, 13, 3, 10, 19, 20, 6, 16, 22, 23, 24, 39, 44, 46 and 50.

Genes identified based on the primers from the sequence of the genome have no characteristic (based on the sequence) of the exported proteins. They are the following sequences: SEQ ID No. 4, 27, 11, 12, 14, 7, 15, 17, 18, 21, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 45, 47, 48 and 49.

Based on the sequence of other organisms such as *E.coli*, it is possible to search in the sequence of the *M. tuberculosis* genome for genes possessing similarities with proteins known to be exported in

other organisms although not possessing an export signal sequence. In this case, fusion with *phoA* is an advantageous protocol for determining if these *M. tuberculosis* sequences encode exported proteins although possessing no consensus signal sequence. It has indeed been possible to clone SEQ ID No. 49, a sequence similar to an *E.coli* gene of the *htrA* family. A fusion of SEQ ID No. 49 with *phoA* leads to the expression and the export of *phoA*. *M. smegmatis* colonies harboring an SEQ ID No. 49 *phoA* fusion on a plasmid pJVED are blue.

SEQ ID No. 49 is therefore considered as an exported protein.

The *phoA* method is therefore useful for detecting, based on the *M. tuberculosis* sequence, genes encoding exported proteins without them encoding sequences which are characteristic of the exported proteins.

Even if a sequence possesses determinants of exported proteins, this does not demonstrate a functional export. The *phoA* system makes it possible to show that the gene suspected really encodes an exported protein. Thus, it was checked that the sequence SEQ ID No. 50 indeed possessed export signals.

TABLE 3

SEQ ID No.	Reference of the corresponding sequence predicted by Cole et al.		Annotation
SEQ ID No. 1	Rv 0203	*	Sequence hydrophobic at the N-terminus
SEQ ID No. 4 SEQ ID No. 27	Rv 2050		No prediction
SEQ ID No. 8 SEQ ID No. 9	Rv 2563	*	Membrane protein
SEQ ID No.	Rv 0072	*	Possible trans-

8G', H'			membrane transport protein of the ABC type
SEQ ID No. 11	Rv 0546c	ML	Protein S-D Lactoyl Glutathione-methyl glyoxal lyase
SEQ ID No. 12	no prediction		not found in *M.tuberculosis* H37rv
SEQ ID No. 13 SEQ ID No. 3 SEQ ID No. 10	Rv 1984c	*	probable precursor cutinase with an N-terminal signal sequence
SEQ ID No. 14 SEQ ID No. 7	no prediction		no prediction
SEQ ID No. 15	with reading frame shift, could be in phase with Rv 2530c		no prediction
SEQ ID No. 17	Rv 1303	ML	no prediction
SEQ ID No. 18	Rv 0199	ML	no prediction
SEQ ID No. 19	Rv 0418	*	site for attachment of prokaryotic membrane lipo-protein, similarity with N-acetyl puromycin acetyl hydrolase
SEQ ID No. 20 SEQ ID No. 6	Rv 3576	*	contains a site for attachment of prokaryotic mem-brane lipoprotein, similarity with a serine/threonine protein kinase
SEQ ID No. 21	Rv 3365c	ML	similarity with a zinc metallo-peptidase

SEQ ID No. 31	not predicted		no prediction
SEQ ID No. 32	Rv 0822c	ML	Existence of a consensus region with the drac family
SEQ ID No. 33	Rv 1044		no prediction
SEQ ID No. 34	not predicted		no prediction
SEQ ID No. 35	Rv 2169c		no prediction
SEQ ID No. 36	Rv 3909	ML	no prediction
SEQ ID No. 37	Rv 2753c		similarity with dihydropricolinate synthases
SEQ ID No. 38	Rv 0175		no prediction
SEQ ID No. 39	Rv 3006	* ML	prediction of lipoprotein signal sequence
SEQ ID No. 40	Rv 0549c		no prediction
SEQ ID No. 41	Rv 2975c being capable of being in phase with Rv 2974c		similarity with substilis protein
SEQ ID No. 42	Rv 2622		similarity with a methyl transferase
SEQ ID No. 43	Rv 3278c	ML	no prediction
SEQ ID No. 44	Rv 0309	*	no prediction
SEQ ID No. 45	Rv 2169c	ML	no prediction
SEQ ID No. 46	Rv 1411c	*	probable lipo-protein with an N-terminal signal sequence
SEQ ID No. 47	Rv 1714		similarity with a gluconate 3-dehydrogenase
SEQ ID No. 48	Rv 0331		similarity with a sulfide dehydro-genase and a sulfide quinone reductase

SEQ ID No. 49	Rv 0983	ML	similarity with a serine protease HtrA
SEQ ID No. 5			
SEQ ID No. 16	Rv 3810	* ML	Surface protein Berthelet et al. 1995
SEQ ID No. 22 SEQ ID No. 23 SEQ ID No. 24	Rv 3763	*	Contains a site for attachment of eukaryotic membrane lipoprotein
SEQ ID No. 50	Rv 0125	*	Active site of serine proteases Possible N-terminal signal sequence

Legend to Table 3:
Correspondence between the sequences according to the invention and the sequences predicted by Cole et al., 1998, Nature, 393, 537-544.

* : Prediction that the protein encoded by the sequence is exported

ML : Prediction of similarity with *M. leprae*.

Example 6:

Characteristics and production of the protein M1C25

The N-terminal end of the protein M1C25 was detected by the *PhoA* system as allowing the export of the fusion protein, necessary for the production of its phosphatase activity.

The DNA sequence encoding the N-terminal end of the protein M1C25 is contained in the sequence SEQ ID No. 20 of the present patent application.

From this primer sequence, the complete gene encoding the protein M1C25 was sought in the *M. tuberculosis* genome (Welcome Trust Foundation, Sanger site).

The Sanger center attributed to M1C25 the names:
Rv 3576,
MTCY06G11.23,
pknM

Sequence SEQ ID No. 29 of the complete M1C25 gene (714 bases):
cf. Figure 29

This gene encodes a protein of 237 AA, having a molecular mass of 25 kDa. This protein is listed in the libraries under the names:
PID:e306716,
SPTREMBL:P96858

Sequence SEQ ID No. 30 of the protein M1C25 (237 amino acids): cf. Figure 30

M1C25 contains a site for attachment to the lipid portion of the prokaryotic membrane lipoproteins (PS00013 Prokaryotic membrane lipoprotein lipid attachment site:
CTGGTCGGTG CGTGCATGCT CGCAGCCGGA TGC).

The function of M1C25 is not clear but it most probably possesses a "serine/threonine protein kinase" activity. Similarities should be noted with the C-terminal moiety of K08G_MYCTU Q11053 Rv1266c (MTCY50.16). Similarities are also found with KY28_MYCTU.

A gene potentially encoding a regulatory protein (PID:e306715, SPTREMBL:P96857, Rv3575c, (MTCY06G11.22c)) is found in 5' of the gene encoding M1C25.

The hydrophobicity profile (Kyte and Doolitle) of M1C25 is represented in Figure 56.

A site of cleavage of the signal sequence is predicted (SignalP V1.1; World Wide Web Prediction Server, Center for Biological Sequence Analysis) between amino acids 31 and 32: AVA-AD. This cleavage

site is behind a conventional "AXA" motif. This prediction is compatible with the hydrophobicity profile. In this potential signal sequence, it is observed that the sequence of the three amino acids LAA is repeated three times.

Cloning of the M1C25 gene for the production of the protein which it encodes:

A pair of primers were synthesized in order to amplify, using the genomic DNA of *M. tuberculosis*, strain H37Rv, the entire sequence encoding the polypeptide M1C25. The amplicon obtained was cloned into an expression vector.

Pairs of primers appropriate for the amplification and the cloning of the sequence encoding M1C25 were synthesized:

- forward primer:

5'-ATAATACCATGGGCAAGCAGCTAGCCGCGC- 3'

- backward primer:

5'-ATTTATAGATCTCTGCTTAGCAACCTTGGCCGCG- 3'

The underlined portion represents the sequences specifically hybridizing with the M1C25 sequence and the 5' ends correspond to restriction sites for the cloning of the resulting amplicon into a cloning and/or expression vector.

A specific vector used for the expression of the polypeptide M1C25 is the vector pQE60 marketed by the company Qiagen, following the protocol and the recommendations proposed by this brand.

The cells used for the cloning are bacteria: *E.coli* XL1-Blue (resistant to tetracycline).

The cells used for the expression are bacteria: *E.coli* M15 (resistant to kanamycin) containing the plasmid pRep4 (M15 pRep4).

The production of the protein MYC25 is illustrated by Figures 57 A and B (bacterial extracts from the *E.coli* M15 strain containing the plasmid pM1C25). The bacterial cultures and the extracts are prepared

according to Sambrook et al. (1989). Analysis of the bacterial extracts is carried out according to the Quiagen instructions (1997).

BIBLIOGRAPHIC REFERENCES

AIDS therapies, 1993, in Mycobacterial infections, ISBN 0-9631698-1-5, pp. 1-11.

Altschul, S.F. et al., 1990, J. Mol. Biol., **215** : 403-410.

Andersen, P. et al., 1991, Infect. Immun., **59** : 1905-1910.

Andersen, P. et al., 1995, J. Immunol., **154** , 3359-3372.

Bange, F.C. et al., A.M. Brown, and W.R. Jacobs JR., 1996, Leucine auxotrophy restricts growth of *Mycobacterium bovis* BCG in macrophages. Infect. Immun., **64**, : 1794-1799.

Barany, F., 1911, Proc. Natl. Acad. Sci. USA, **88** : 189-193.

Bates, J. et al., 1986, Am. Rev. Respir. Dis., **134** : 415-417.

Bates, J. 1979, Chest. **76**(Suppl.):757-763.

Bates, J. et al.. 1986. Am. Rev. Respir. Dis. **134** : 415-417.

Berthet, F.X., J. Rauzier, E.M. Lim, W. Philipp, B. Gicquel, and D. Portnoï, 1995, Characterization of the *M. tuberculosis erp* gene encoding a potential cell surface protein with repetitive structures. Microbiology. **In press**

Borremans, M. et al., 1989, Biochemistry, 7 : 3123-3130.

Bouvet, E. 1994, Rev. Fr. Lab. **273** : 53-56.

Brockman, R.W. and Heppel L.A., 1968, On the localization of alkaline phosphatase and cyclic phosphodiesterase in *Escherichia coli*, Biochemistry, **7** : 2554-2561.

Burg, J.L. et al., 1996, Mol. and Cell. Probes, **10** : 257-271.

Chevrier, D. et al., 1993, Mol. and Cell. Probes, **7** : 187-197.

Clemens, D.L., 1996, Characterization of the Mycobacterium tuberculosis phagosome, Trends Microbiol., **4** : 113-118.

Chu, B.C.F. et al., 1986, Nucleic Acids Res., **14** : 5591-5603.

Clemens, D.L. and Horwitz M. A., 1995, Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited, J. Exp. Med., **181** : 257-270.

Colignon J.E., 1996, Immumologic studies in humans. Measurement of proliferative responses of culturered lymphocytes. Current Protocols in Immunology, NIH, **2**, Section II.

Daniel, T.M. et al. 1987, Am. Rev. Respir. Dis., **135** : 1137-1151).

Dellagostin, O.A., Esposito G., Eales L.-J., Dale J.W. and. McFadden J.J., 1995, Activity of mycobacterial promoters during intracellular and extracellular growth. Microbiol., **141** : 2123-2130.

Drake, T.A. et al. 1987. J. Clin. Mocrobiol. **25** : 1442-1445.

Dramsi et al., 1997, Infection and Immunity, **65**, 5 : 1615-1625.

Duck, P. et al., 1990, Biotechniques, **9** : 142-147.

Erlich, H.A. 1989. In PCR Technology. Principles and Applications for DNA Amplification. New York: Stockton Press.

Felgner et al., 1987, Proc. Natl. Acad. Sci., **84** : 7413.

Fraley et al., 1980, J. Biol. Chem., **255** : 10431.

Gaillard, J.L., Berche P., Frehel C., Gouin E. and Cossart P., 1991, Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from Gram-positive cocci, Cell., **65** : 1127-1141.

Garbe, T., Harris D., Vordermeir M., Lathigra R., Ivanyi J. and Young D., 1993, Expression of the *Mycobacterium tuberculosis* 19-kilodalton antigen in

Mycobacterium smegmastis: immunological analysis and evidence of glycosylation. Infect. Immun., **61** : 260-267.

Guateli, J.C. et al., 1990, Proc. Natl. Acad. Sci. USA, **87** : 1874-1878.

Harboe et al., 1996, Infect. Immun., **64**, 16-22.

Herrmann, J.L., O'Gaora P., Gallagher A., Thole J.E.R. and Young D.B., 1996, Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*, EMBO J. **15** : 3547-3554.

Houbenweyl, 1974, in Meuthode der Organischen Chemie, E. Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart.

Huygen, K. et al., 1996, Nature Medicine, **2**(8) : 893-898.

Innis, M.A. et al., 1990. in PCR Protocols. A guide to Methods and Applications. San Diego: Academic Press.

Isberg, R.R., Voorhis D.L. and Falkow S., 1987, Identification of invasin: a protein that allows enteric bacteria to penetrate cultured mammalian cells, Cell, **50** : 769-778.

Jacobs, W.R. et al., 1991. Construction of mycobacterial genomic libraries in shuttle cosmids. Genetic Systems for Mycobacteria, Methods in Enzymology, **204** : 537-555.

Jacobs, W.R. et al., 1993, Science, **260** : 819-822.

Kaneda, et al., 1989, Science, **243**:375.

Kiehn, T.E., et al. 1987. J. Clin. Microbiol. **25** : 1551-1552.

Kievitis, T. et al., 1991, J. Virol. Methods, **35** : 273-286.

Kohler, G. et al., 1975, Nature, **256**(5517):495-497.

Kwoh, D.Y. et al., 1989, Proc. Natl. Acad. Sci. USA, **86** : 1173-1177.

Landegren, U. et al., 1988, Science, **241**, : 1077-1080.

Lang, T. and Antoine J.-C., 1991, Localization of MHC classII molecules in murine bone marrow-derived macrophages. Immunology, **72** : 199-205.

Lee, B.Y and Horwitz M.A., 1995, Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*, J. Clin. Invest., **96** : 245-249.

Lim, E.M., Rauzier J., Timm J., Torrea G., Murray A., Gicquel B. and Portnoï D., 1995, Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins, using *phoA* gene fusions, J. Bacteriol., **177** : 59-65.

Lizardi, P.M. et al., 1988, Bio/technology, **6** : 1197-1202.

Mahan, M.J. et al., 1993. Selection of bacterial virulence genes that are specifically induced in host tissues, Science, **259** : 686-688.

Manoil L., Mekolanos J.J. and Beckwith J., J. Bacteriol., 1990, **172**, 515-518.

Matthew, J.A. et al., 1988, Anal. Biochem., **169** : 1-25.

Merrifield, R.D., 1966, J. Am. Chem. Soc., **88**(21) : 5051-5052.

Midoux, 1993, Nucleic Acids Research, **21** : 871-878/

Miele, E.A. et al., 1983, J. Mol. Biol., **171** : 281-295.

Minton, N.P., 1984, Gene, **31**, 269-273.

Montgomery et al., 1993, DNA Cell Biol., **12** : 777-783.

Navarre, W.W. et al., 1994, Molecular Microbiologie, **14**(1) : 115-121.

Navarre, W.W. et al., 1996, J. of Bacteriology, 178, **2** : 441-446.

Pagano et al., 1967, J. Virol., **1** : 891.

Pastore, 1994, Circulation, **90** : I-517.

Patel, et al. 1990. J. Clin. Microbiol. : 513-518.

Prentki, B. et Krish H. M., 1984, Gene, **29** : 303-313.

Pettersson R., Nordfelth J., Dubinina E., Bergman T., Gustafsson M., Magnusson K.E. and Wolf-Watz H., 1996, Modulation of virulence factor expression by pathogen target cell contact. Science., **273** : 1231-1233.

Plum, G. and Clark-Curtiss J.E., 1994, Induction of *Mycobacterium avium* gene expression following phagocytosis by human macrophages. Infect. Immun., **62** : 476-483.

Roberts, M.C., et al. 1987. J. Clin. Microbiol. **25**, : 1239-1243.

Rolfs, A. et al. 1991. In PCR Topics. Usage of Polymerase Chain reaction in Genetic and Infectious Disease. Berlin: Springer-Verlag.

Sambrook, J. et al. 1989. In Molecular cloning : A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sanchez-Pescador, R., 1988, J. Clin. Microbiol., **26**(10),:1934-1938.

Schneewind, O. et al., 1995, Science, **268** : 103-106.

Segev D., 1992, in "Non-radioactive Labeling and Detection of Biomolecules". Kessler C. Springer Verlag, Berlin, New-York, 197-205.

Servant, P. and Mazodier P., 1995, Characterization of *Streptomyces albus* 18-kilodalton heat shock-responsive protein. J. Bacteriol., **177** : 2998-3003.

Shiver, J.W., 1995, in Vaccines 1995, eds Chanock, R.M. Brown, F. Ginsberg, H.S. & Norrby, E.), pp.95-98, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sorensen et al., 1995, Infect. Immun., **63**, 1710-1717.

Stone, B.B. et al., 1996, Mol. and Cell. Probes, **10** : 359-370.

Stover, C.K., Bansal G.P., Hanson M.S., Burlein S.R., Palaszynski S.R., Young J.F., Koenig S., Young D.B., Sadziene A. and Barbour A.G., 1993, Protective immunity elecited by recombinant Bacille Calmette-Guerin (BCG) expressing outer surface protien A (OspA) lipoprotein: a candidate Lyme disease vaccine. J. Exp. Med., **178** : 197-209.

Sturgill-Koszycki, S., Schlesinger P.H., Chakroborty P., Haddix P.L., Collins H.L., Fok A.K., Allen R.D., Gluck S.L., Heuser J. and Russell D.G., 1994, Lack of acidification of *Mycobacterium* phagosomes by exclusion of the vesicular proton-ATPase. Science., **263** : 678-681.

Tascon, R.E. et al.., 1996, Nature Medicine, 2(8):888-892.

Technique for assembling oligonucleotides, 1983, Proc. Natl. Acad. Sci. USA, **80** : 7461-7465.

Technique for beta-cyanethylphosphoramidites, 1986, Bioorganic Chem., **4**: 274-325.

Thierry, D. et al., 1990, Nucl. Acid Res., **18** : 188.

Timm, J., Perilli M.G., Duez C., Trias J., Orefici G., Fattorini L., Amicosante G., Oratore A., Boris B., Frere J.M., Pugsley A.P. and Gicquel B., 1994, Transcription and expression analysis, using *lacZ* and *phoA* gene fusions, of *Mycobacterium fortuitum* B-lactamase genes cloned from a natural isolate and a high-level B-lactamase producer. Mol. Microbiol., **12** : 491-504.

Tuberculosis Prevention Trial, 1980, Mendis, "Trial of BCG vaccines in South India for Tuberculosis Infection", Indian Journal of Medical research, **1972 (Suppl.)**: 1-74.

Urdea, M.S. et al., 1991, Nucleic Acids Symp. Ser., **24** : 197-200.

Urdea, M.S., 1988, Nucleic Acids Research, **11** : 4937-4957.

Verbon, A., Hartskeerl R.A., Schuitema A., Kolk A.H., Young D.B. and Lathigra R., 1992, The 14,000-molecular-weight antigen of *Mycobacterium tuberculosis* is related to the alpha-crystallin family of low-molecular-weight heat shock proteins. J Bacteriol., **174** : 1352-1359.

Walker, G.T. et al., 1992, Nucleic Acids Res., **20** : 1691-1696.

Walker, G.T. et al., 1992, Proc. Natl. Acad. Sci. USA, **89** : 392-396.

Wiker, H.G. et al., 1992, Microbiol. Rev., **56** : 648-661.

Yamaguchi, R. et al., 1989, Infect Immun., **57** : 283-288;

Xu, S., Cooper A., Sturgill-Koszycki S., van Heyningen T., Chatterjee D., Orme I., Allen P. and Russel D.G.,

1994, Intracellular trafficking in *Mycobacterium tuberculosis* and *Mycobacterium avium*-infected macrophages, J. Immuno., **153**: 2568-2578.

Young, D.B. et al., 1992, Mol. Microbiol., **6** : 133-145.

Yuen, L.K.W. et al., 1993, J. Clin. Microbiol., **31** : 1615-1618.